

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:45:45 ; Search time 13.25 Seconds
(without alignments)

1253.522 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2368
Sequence: 1 MKIFKCYFKHTLQCKRFLIF.....DWITLPSEKLFMDNRNLTTTS 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	863.5	36.1	428	1 G6NT_HUMAN	Q02742 homo sapien
2	848.5	35.5	428	1 G6NT_MOUSE	Q09324 mus musculu
3	840	35.2	427	1 G6NT_BOVIN	Q92180 bos taurus
4	730.5	30.6	400	1 BGIB_MOUSE	P97402 mus musculu
5	711	29.8	400	1 BGIB_HUMAN	Q06430 homo sapien
6	715	4.9	895	1 YAS5_METJA	Q58454 metanococ
7	108.5	4.5	661	1 WHI1_YEAST	P34761 saccharomyc
8	102.5	4.3	654	1 BFR2_HUMAN	Q01742 homo sapien
9	102.5	4.3	821	1 FGR2_HUMAN	P21802 homo sapien
10	102.5	4.3	821	1 FGR2_MOUSE	Q09604 ureaplasma
11	101.5	4.2	1442	1 DPO3_UREPA	Q99604 ureaplasma
12	101.5	4.2	2054	1 YCF2_PINTH	P41653 pinus thunb
13	101.5	4.2	2280	1 YCF2_TOBAC	P09976 nicotiana t
14	100.5	4.2	2607	1 BACH_BACLI	O68007 bacillus l
15	99.5	4.2	666	1 VOO1_VACCC	P21093 vaccinia v
16	99.5	4.2	823	1 CEK3_CHICK	P18461 gallus gall
17	99	4.1	398	1 A23D_DROME	Q24093 drosophila
18	97.5	4.1	934	1 SYL1_SUISO	P58176 sulfolobus
19	97.5	4.1	1314	1 TERY_CLOTE	P04858 clostridium
20	97	4.1	340	1 YM2_ARCFU	O28062 archaeoglob
21	97	4.1	440	1 GAAP_HUMAN	Q00591 homo sapien
22	97	4.1	1036	1 YOB6_CAEEL	Q09559 caenorhabd
23	95.5	4.0	936	1 MSH4_HUMAN	Q15457 homo sapien
24	95	4.0	701	1 SYGB_HELPJ	Q92489 helicobacte
25	94	3.9	364	1 DP3B_MYCGE	P47247 mycoplasma
26	94	3.9	454	1 DNAA_BUCAP	P29434 buchnera ap
27	94	3.9	1344	1 IFN3_MOUSE	P23116 mus musculu
28	94	3.9	1675	1 CLH1_HUMAN	Q00610 homo sapien
29	94	3.9	1675	1 CLH1_BOVIN	P49951 bos taurus
30	94	3.9	1674	1 CLH1_RAT	P11442 rattus norv
31	94	3.9	5255	1 BACA_BACLI	O68006 b bacillrac
32	93.5	3.9	884	1 SECA_PORPU	P51381 porphyra pu
33	93.5	3.9	2787	1 TEL1_YEAST	P38110 saccharomyc

34	93	3.9	503	1 CPV1_SHEEP	O35278 ovis aries
35	93	3.9	520	1 Y308_METJA	O60270 methanococ
36	92	3.9	908	1 DPO1_BORBU	O51498 borrelia bu
37	92	3.9	1103	1 CYGF_BOVIN	O02740 bos taurus
38	92	3.9	1382	1 IF3A_HUMAN	O14152 homo sapien
39	92	3.9	2022	1 ANM1_ONCVO	P21245 onchocerca
40	91.5	3.8	720	1 SPOT_MYCGE	P47520 mycoplasma
41	91.5	3.8	757	1 FRS1_YEAST	P38903 saccharomyc
42	91.5	3.8	1956	1 ARX1_PLARA	O04956 plasmodium
43	91	3.8	344	1 Y613_METJA	O68030 methanococ
44	91	3.8	458	1 YC13_EUGGR	P05728 euglena gra
45	91	3.8	490	1 MET3_SCHPO	P78937 schistosach

ALIGNMENTS

```

RESULT 1
ID G6NT_HUMAN STANDARD: PRT: 428 AA.
AC Q02742;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-
DE ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME)
DE (CORE2-GLCNAC-TRANSFERASE) (C2GNT) (CORE 2 GNT).
GN GCNT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93028457; PubMed=1329093;
RA Bierhuizen M.F.A., Fukuda M.;
RT "Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta
RT 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6-GlcNAc transferase by gene
RT transfer into CHO cells expressing polyoma large tumor antigen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9326-9330(1992).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96078409; PubMed=759796;
RA Bierhuizen M.F.A., Maemura K., Kudo S., Fukuda M.;
RT "Genomic organization of core 2 and I branching beta-1,6-N-
RT acetylglucosaminyltransferases. Implication for evolution of the
RT beta-1,6-N-acetylglucosaminyltransferase gene family.";
RL Glycobiology 5:417-425(1995).
CC - FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
CC - CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-
CC GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYL-R -> UDP + BETA-D-
CC GALACTOSYL-1,3-N-ACETYL-D-GLUCOSAMINYL-R.
CC - PATHWAY: GLYCOSYLATION.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, GOLGI.
CC - TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ACTIVATED T-LYMPHOCYTES
CC AND MYELOID CELLS.
CC - SIMILARITY: TO I-BRANCHING ENZYME (IGNT).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M97347; AAA35919.1;
CC EMBL: L41415; AAA96661.1;
CC PIR: A46293; A46293.
CC MIM: 600391;
CC InterPro: IPR003406; Branch.

```

DR Pfam: PF02485; Branch: 1.
 Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 FT DOMAIN 1
 FT TRANSMEM 10 32
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 33 428
 FT CARBOHYD 52 52
 FT CARBOHYD 58 58
 FT CARBOHYD 95 95
 FT CARBOHYD 95 95
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 428 AA: 49784 MW, 769ECB51DA00A60 CRC64;
 SQ

Query Match 36.1%; Score 863.5; DB 1; Length 428;
 Best Local Similarity 43.1%; Pred. No. 2,7e-56;
 Matches 187; Conservative 72; Mismatches 122; Indels 43; Gaps 12;

QY 28 LKLLNVRRLP--POKDIYLV-EVSLSTSPVNRNRYTHVDE--VRY-----EV 71
 DB 1 MRLTLRRRLFSPTKRYEYMLVSLITFVYLR--IHQKPEFVSRLHLELAGENPSSDI 57
 QY 72 NCSGIYEOEPLERIGK-----SLFIRRDITLEDVVAMTSDCDIYQTLRGYAKLYS 125
 DB 58 NCTKVLQGDVNEIOKVKLEILTVKFKRP--RWTPDDYINMTSDCSSIRKRYIVEPLS 115
 QY 126 KEKSPFIAYSLVHKDALMWERLHAIFYNQHNYCIHYDRKAPDTEFVAMNNAKCFSN 185
 DB 116 KEKSPFIAYSLVHKDALMWERLHAIFYNQHNYCIHYDRKAPDTEFVAMNNAKCFSN 175
 QY 186 IFIASKLEAVEYAHISRLQADNLCLSDLKSSIQMKYVINLCGDPFLKSNFELYSELK 245
 DB 176 VFAASRLSEYVYASWSRVQADNLCKKDYANSAWKYILINLCGDEPKLTLEIYRKL 235
 QY 246 LINGANMLETVKPNKSKLERFTYHHLRVPYEV--KLPRTNISKAPPHNIQIFVGSAA 303
 DB 236 LINGANMLETVKPNKSKLERFTYHHLRVPYEV--KLPRTNISKAPPHNIQIFVGSAA 287
 QY 304 YVLSQAFKYIFNNSIVODFFAMSKDYSPDEHFWATLIRVPGIGETISASQ-DVSD 362
 DB 288 YVLSQAFKYIFNNSIVODFFAMSKDYSPDEHFWATLIRVPGIGETISASQ-DVSD 347
 QY 363 QSKTRLVKKNYVEGFF-----YPSCTGSHLRVSVCIYGAELRWLIKDHMPANKFDSKYD 417
 DB 348 QAVAFEVKQYREGDVSNGAPRPPCDGVHVSVCIFGAGDLNMMKRLHNFANFEDVVD 407
 QY 418 PLIKCLAEKLEEQ 431
 DB 408 LFAIQCLDEHLRRK 421

RESULT 2
 GENT_MOUSE STANDARD: PRT: 428 AA.
 ID GENT_MOUSE
 AC 009324;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BETA-1,3-GALACTOSYL-O-GLYCOSYL-GLYCOPROTEIN BETA-1,6-N-ACETYLGLUCOSAMINYLTANSFERASE (EC 2.4.1.102) (CORE 2 BRANCHING ENZYME).
 DE (CORE2-GLCNAC-TRANSFERASE) (CGENT).
 GN GCN1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DA/2;
 RA Warren C.E., Smockler D.S., Dennis J.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FORMS CRITICAL BRANCHES IN O-GLYCANS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL-1,3-N-ACETYL-D-GALACTOSAMINYLT-R = UDP + BETA-D-

CC GALACTOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYLT-1,6)-N-ACETYL-D-
 CC GALACTOSAMINYLT-R.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC -1- SIMILARITY: TO I-BRANCHING ENZYME (INT).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U19265; AAA60948.1; -
 DR MGI: 956767; Gcn1.
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch: 1.
 DR Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack; Glycoprotein.
 FT DOMAIN 1
 FT TRANSMEM 10 32
 FT SIGNAL-ANCHOR (POTENTIAL).
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 33 428
 FT CARBOHYD 58 58
 FT CARBOHYD 95 95
 FT CARBOHYD 95 95
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 428 AA: 49849 MW, F3FED2517827AAE1 CRC64;
 SQ

Query Match 35.5%; Score 848.5; DB 1; Length 428;
 Best Local Similarity 42.4%; Pred. No. 3.4e-55;
 Matches 183; Conservative 72; Mismatches 138; Indels 39; Gaps 11;

QY 28 LKLLNVRRLP--POKDIY-LVEYSLSTSPVNRNRYTHVDE--VRY-----EV 71
 DB 1 MRLTLRRRLFSPTKRYEYMLVSLITFVYLR--IHQKPEFVSRLHLELAGDDPYSNV 57
 QY 72 NCSGIYEOEPLERIGK-----SLFIRRDITLEDVVAMTSDCDIYQTLRGYAKLYS 125
 DB 58 NCTKVLQGDVNEIOKVKLEILTVKFKRP--RWTPDDYINMTSDCSSIRKRYIVEPLT 115
 QY 126 KEKSPFIAYSLVHKDALMWERLHAIFYNQHNYCIHYDRKAPDTEFVAMNNAKCFSN 185
 DB 116 KEKSPFIAYSLVHKDALMWERLHAIFYNQHNYCIHYDRKAPDTEFVAMNNAKCFSN 175
 QY 186 IFIASKLEAVEYAHISRLQADNLCLSDLKSSIQMKYVINLCGDPFLKSNFELYSELK 245
 DB 176 VFAASRLSEYVYASWSRVQADNLCKKDYANSAWKYILINLCGDEPKLTLEIYRKL 235
 QY 246 LINGANMLETVKPNKSKLERFTYHHLRVPYEVK-LPRTNISKAPPHNIQIFVGSAYF 305
 DB 236 LINGANMLETVKPNKSKLERFTYHHLRVPYEVK-LPRTNISKAPPHNIQIFVGSAYF 289
 QY 306 YVLSQAFKYIFNNSIVODFFAMSKDYSPDEHFWATLIRVPGIGETISASQ-DVSD 364
 DB 290 YVLSQAFKYIFNNSIVODFFAMSKDYSPDEHFWATLIRVPGIGETISASQ-DVSD 349
 QY 365 KTRLVKKNYVEGFF-----YPSCTGSHLRVSVCIYGAELRWLIKDHMPANKFDSKYD 419
 DB 350 IAFEVKQYREGDVSNGAPRPPCDGVHVSVCIFGAGDLNMMKRLHNFANFEDVVD 409
 QY 420 PLIKCLAEKLEEQ 431
 DB 410 AIOCLDEHLRRK 421

RESULT 3
 GENT_BOVIN STANDARD: PRT: 427 AA.
 ID GENT_BOVIN
 AC 092180;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

QY	365	KITLVNWNVEE	-----	YPSCTSHLRVCIQAEMLMDGWEFNKRSYVDI	419
DB	350	IAFVAKMVFEDDYSKGNPIPCRS-VHRSVCVCGAGDMLMLVHHLFNNKRDITDLE			408
QY	420	LKCLAEKLEEQ	431		
DB	409	AIQCLDEHRLRK	420		
RESULT	4				
ID	SGIB_MOUSE	STANDARD:	PRT:	400 AA.	
AC	P97402;				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	N-ACETYLGLYCOSAMINE BETA-1,6-N-ACETYLGLYCOSAMINYLTANSFERASE				
DE	(EC 2.4.1.150) (N-ACETYLGLYCOSAMINYLTANSFERASE) (I-BRANCHING ENZYME)				
DE	(IGNT) (LARGE I ANTIGEN-FORMING BETA-1,6-N-				
DE	ACETYLGLYCOSAMINYLTANSFERASE).				
GN	GCNT2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID:10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97280061; PubMed=9134435;				
RT	Magnet A.D., Fukuda M.				
RT	"Expression of the large I antigen forming beta-1,6-N-				
RT	acetylglucosaminyltransferase in various tissues of adult mice."				
RL	Glycobiology 7:285-295(1997).				
CC	- FUNCTION: BRANCHING ENZYME THAT CONVERTS LINEAR INTO BRANCHED				
CC	POLY-N-ACETYLGLYCOSAMINOGLYCANS. INTRODUCES THE BLOOD GROUP I				
CC	ANTIGEN DURING EMBRYONIC DEVELOPMENT. IT IS CLOSELY ASSOCIATED				
CC	WITH THE DEVELOPMENT AND MATURATION OF ERYTHROID CELLS.				
CC	- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + BETA-D-GALACTOSYL				
CC	-1,4-N-ACETYL-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-				
CC	GLUCOSAMINYL-1,6-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.				
CC	- PATHWAY: GLYCOSYLATION				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.				
CC	- SIMILARITY: TO CORE 2 BRANCHING ENZYME (CGNT).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed, usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: U6182; AAB93621.1; .				
DR	MGI: MGI:1100870; Gcmt2.				
DR	InterPro: IPR003406; Branch.				
DR	Pfam: PF02485; Branch; 1..				
KW	Transferase; Glycosyltransferase; Transmembrane; Signal-anchor;				
KW	Golgi stack; Glycoprotein.				
FT	DOMAIN	1			
FT	TRANSMEM	7	25		
FT		6			
FT					
FT	DOMAIN	26	400		
FT	CARBOHYD	37	37		
FT	CARBOHYD	255	255		
FT	CARBOHYD	314	314		
FT	CARBOHYD	388	388		
FT	SEQUENCE	400 AA;	4550 MW;		
FT			193FF882EE3B9691 CRC64;		
Query Match	30.6%	Score 730.5;	DB 1;	Length 400;	
Best Local Similarity	40.0%	Pred. No. 1.5e-46;			
Matches 156;	Conservative 51;	Mismatches 160;	Indels 23;	Gaps 6	

RN [1]
 RA SEQUENCE FROM N.A.
 RP TISSUE-Neonatal brain stem;
 RC MEDLINE=9036077; PubMed=1697263;
 RX Dionne C.A., Crumley G.R., Bellot F., Kaplow J.M., Searfoss G.,
 RA Rida M., Burgess W.H., Jaye M., Schlessinger J.;
 RT "Cloning and expression of two distinct high-affinity receptors
 RT cross-reacting with acidic and basic fibroblast growth factors.";
 RL EMBO J. 9:2685-2692(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91045961; PubMed=2172978;
 RA Housseaint E., Blaquet P.R., Champion-Arnaud P., Gesnel M.C.,
 RA Torriglia A., Courtois Y., Breathnach R.;
 RT "Related fibroblast growth factor receptor genes exist in the human
 RT genome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8180-8184(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91274356; PubMed=1647213;
 RA Seno M., Sasada R., Watanabe T., Ishimaru K., Igarashi K.;
 RT "Two CDNA's encoding novel human FGF receptor.";
 RL Biochim. Biophys. Acta 1089:244-246(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Stomach cancer;
 RX MEDLINE=90332706; PubMed=2377625;
 RA Hattori Y., Odagiri H., Nakatani H., Miyagawa K., Naito K.,
 RA Sakamoto H., Katoh O., Yoshida T., Sugimura T., Terada M.;
 RT "K-sam, an amplified gene in stomach cancer, is a member of the
 RT heparin-binding growth factor receptor genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5983-5987(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92212948; PubMed=1313574;
 RA Katoh M., Hattori Y., Sasaki H., Tanaka M., Sugano K., Yazaki Y.,
 RA Sugimura T., Terada M.;
 RT "K-sam gene encodes secreted as well as transmembrane receptor
 RT tyrosine kinase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:2960-2964(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=93016048; PubMed=1400433;
 RA Dell K.R., Williams L.T.;
 RT "A novel form of fibroblast growth factor receptor 2. Alternative
 RT splicing of the third immunoglobulin-like domain confers ligand
 RT binding specificity.";
 RL J. Biol. Chem. 267:21225-21229(1992).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Blood;
 RA Steinberger D., Mueller U.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP VARIANTS CS HIS-340; ARG-342; SER-342; TYR-342 AND CYS-354.
 RX MEDLINE=95078932; PubMed=7987400;
 RA Reardon W., Winter R.M., Rutland P., Pulley L.J., Jones B.M.,
 RA Malcolm S.;
 RT "Mutations in the fibroblast growth factor receptor 2 gene cause
 RT Crouzon syndrome.";
 RL Nat. Genet. 8:98-103(1994).
 RN [9]
 RP VARIANTS CS CYS-328 AND CYS-347, AND VARIANT JWS GLY-344.
 RX MEDLINE=95179174; PubMed=7874170;
 RA Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J.,
 RA Charnas L.R., Jackson C.E., Jaye M.;
 RT "Jackson-Weiss and Crouzon syndromes are allelic with mutations in
 RT fibroblast growth factor receptor 2.";
 RL Nat. Genet. 8:275-279(1994).
 RN [10]
 RP VARIANTS CS.
 RX MEDLINE=95384152; PubMed=7655462;
 RA Oldridge M., Wilkie A.O.M., Stanley S.F., Poole M.D., Pulley L.J.,
 RA Rutland P., Hockley A.D., Wake M.J.C., Goldin J.H., Winter R.M.,
 RA Reardon W., Malcolm S.;
 RT "Mutations in the third immunoglobulin domain of the fibroblast growth
 RT factor receptor-2 gene in Crouzon syndrome.";
 RL Hum. Mol. Genet. 4:1077-1082(1995).
 RN [11]
 RP VARIANT PS ALA-321.
 RX MEDLINE=95235551; PubMed=7719333;
 RA Lafenite E., Wei M.H., Bonaventure J., Munnich A., Le Merrer M.,
 RA Renier D.;
 RT "FGFR2 mutations in Pfeiffer syndrome.";
 RL Nat. Genet. 9:108-108(1995).
 RN [12]
 RP VARIANTS AS TRP-252 AND ARG-253.
 RX MEDLINE=95235562; PubMed=7719344;
 RA Wilkie A.O.M., Stanley S.F., Oldridge M., Poole M.D., Ashworth G.J.,
 RA Hockley A.D., Hayward R.D., David D.J., Pulley L.J., Rutland P.,
 RA Malcolm S., Winter R.M., Reardon W.;
 RT "Apeft syndrome results from localized mutations of FGFR2 and is
 RT allelic with Crouzon syndrome.";
 RL Nat. Genet. 9:165-172(1995).
 RN [13]
 RP VARIANTS PS PRO-341; ARG-342 AND TYR-342.
 RX MEDLINE=95235563; PubMed=7719345;
 RA Rutland P., Pulley L.J., Reardon W., Baralster M., Hayward R.,
 RA Poole M.D., Wilkie A.O.M.;
 RT "Identical mutations in the FGFR2 gene cause both Pfeiffer and
 RT Crouzon syndrome phenotypes.";
 RL Nat. Genet. 9:173-176(1995).
 RN [14]
 RP VARIANTS CS CYS/PS.
 RX MEDLINE=96203627; PubMed=8644708;
 RA Meyers G.A., Day D., Goldberg R., Daencl D.L., Przybyla K.A.,
 RA Abrams L.J., Graham J.M. Jr., Feingold M., Moeschler J.B.,
 RA Ramsley E., Scott A.F., Jabs E.W.;
 RT "FGFR2 exon 11a and 11c mutations in Crouzon, Jackson-Weiss, and
 RT Pfeiffer syndromes: evidence for missense changes, insertions, and a
 RT deletion due to alternative RNA splicing.";
 RL Am. J. Hum. Genet. 58:491-498(1996).
 RN [15]
 RP VARIANTS CS CYS-105; GLU-338; CYS-351 AND ARG-384.
 RX MEDLINE=97101656; PubMed=8946174;
 RA Pulley L.J., Reardon W., Wilkes D., Rutland P., Jones B.M.,
 RA Hayward R., Hall C.M., Brunton L., Chun N., Lammer E., Malcolm S.,
 RA Winter R.M.;
 RT "Spectrum of craniosynostosis phenotypes associated with novel
 RT mutations at the fibroblast growth factor receptor 2 locus.";
 RL Eur. J. Hum. Genet. 4:283-291(1996).
 RN [16]
 RP VARIANTS CS ILE-331; ASP-336--ALA-337 DUPL AND TRP-356--THR-358 DEL.
 RX MEDLINE=97114301; PubMed=8956050;
 RA Steinberger D., Mulliken J.B., Mueller U.;
 RT "Crouzon syndrome: previously unrecognized deletion, duplication, and
 RT point mutation within FGFR2 gene.";
 RL Hum. Mutat. 8:386-390(1996).
 RN [17]
 RP VARIANT PS CYS-290.
 RX MEDLINE=97295073; PubMed=9150725;
 RA Tartaglia M., Valeri S., Velardi F., di Rocco C., Battaglia P.A.;
 RT "Trp90Cys mutation in exon 11a of the fibroblast growth factor
 RT receptor 2 (FGFR2) gene is associated with Pfeiffer syndrome.";
 RL Hum. Genet. 99:602-606(1997).
 RN [18]
 RP VARIANT CS L-252, VARIANT AS F-252, AND VARIANT PS F-252--S-253.
 RX MEDLINE=97156222; PubMed=9002682;
 RA Oldridge M., Lunt P.W., Zackai E.H., McDonald-Mcginn D.M., Muenke M.,
 RA Woloney D.M., Twigg S.R.F., Heath J.K., Howard T.D., Hoganson G.,
 RA Gagnon D.M., Jabs E.W., Wilkie A.O.M.;
 RT "Genotype-phenotype correlation for nucleotide substitutions in the
 RT 1911-1911 linker of FGFR2.";
 RL Hum. Mol. Genet. 6:137-143(1997).

RN [19]
 RP VARIANT CS GRU-292.
 RX MEDLINE-97297373; PubMed-9152842.
 RA Steinberger D., Collmann H., Schmalenberger B., Mueller U.;
 RT "A novel mutation (a885g) in exon 5 of FGFR2 in members of a family
 RT with Crouzon phenotype and plagiocephaly.";
 RL J. Med. Genet. 34:420-422(1997).
 RN [20]
 RP VARIANTS CS VAL-276; CYS-301 AND SER-314.
 RX MEDLINE-98180879; PubMed-9521581.
 RA Steinberger D., Vriend G., Mulliken J.B., Mueller U.;
 RT "The mutations in FGFR2-associated craniosynostoses are clustered in
 RT five structural elements of immunoglobulin-like domain III of the
 RT receptor.";
 RL Hum. Genet. 102:145-150(1998).
 RN [21]
 RP VARIANTS AS TRP-252 AND ARG-253.
 RX MEDLINE-98112406; PubMed-9452027;
 RA Tsai F.-J., Hwu W.-L., Lin S.-P., Chang J.-G., Wang T.-R., Tsai C.-H.;
 RT "Two common mutations 934C to G and 937C to G of fibroblast growth
 RT factor receptor 2 (FGFR2) gene in Chinese patients with Apert
 RT syndrome.";
 RL Hum. Mutat. Suppl. 1:S18-S19(1998).
 RN [22]
 RP VARIANT PS CYS-351.
 RX MEDLINE-98356420; PubMed-9693549;
 RA Mathijssen I.M., Vaandrager J.M., Hoogeboom A.J.,
 RA Hesselting-Janssen A.L., van den Ouweland A.M.W.;
 RT "Pfeiffer's syndrome resulting from an S351C mutation in the
 RT fibroblast growth factor receptor-2 gene.";
 RL J. Craniofac. Surg. 9:207-209(1998).
 CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH
 CC FACTORS.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DISEASE: DEFECTS IN FGFR2 ARE A CAUSE OF CROUNSON SYNDROME (CS),
 CC ALSO CALLED CRANIOFACIAL DYSOSTOSIS TYPE I (CFDI). CHARACTERIZED
 CC BY CRANIOSYNOSTOSIS (PREMATURE FUSION OF THE SKULL SUTURES),
 CC HYPERTELORISM, EXOPHTHALMOS AND EXTERNAL STRABISMUS. PAROC-BEAKED
 CC NOSE, SHORT UPPER LIP, HYPOPLASTIC MAXILLA, AND A RELATIVE
 CC MANDIBULAR PROGNATHISM.
 CC
 Query Match 4.3%; Score 102.5; DB 1; Length 821;
 Best Local Similarity 24.0%; Pred. No. 5.1;
 Matches 78; Conservative 35; Mismatches 109; Indels 103; Gaps 18;

FGFR2_MOUSE
 ID FGFR2_MOUSE STANDARD: PRT; 821 AA.
 AC P21803; 061342;
 DT 01-NOV-1991 (Rel. 18, Created)
 DT 01-NOV-1991 (Rel. 35, Last sequence update)
 DT 20-APR-2001 (Rel. 40, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR RECEPTOR 2 PRECURSOR (FGFR-2)
 DE (KERATINOCYTE GROWTH FACTOR RECEPTOR).
 OS FGFR2 OR EGF1 OR BEK.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Brain, and Liver;
 RX MEDLINE-92228773; PubMed-1373495;
 RA Mansukhani A., Dell'Era P., Moscatelli D., Kornbluth S.,
 RA Hanafusa H., Basilico C.;
 RT "Characterization of the murine BEK fibroblast growth factor (FGF)
 RT receptor: activation by three members of the Fgf family and
 RT requirement for heparin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3305-3309(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RX MEDLINE-91093977; PubMed-1846048;
 RA Miki T., Fleming T.P., Bottaro D.P., Rubin J.S., Ron D.,
 RA Aaronson S.A.;
 RT "Expression cDNA cloning of the KGF receptor by creation of a
 RT transforming autoclone loop.";
 RL Science 251:72-75(1991).
 RN [3]
 RP SEQUENCE OF 477-821 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE-89219016; PubMed-2468999;
 RA Kornbluth S., Paulson K.E., Hanafusa H.;
 RT "Novel tyrosine kinase identified by phosphotyrosine antibody
 RT screening of cDNA libraries.";
 RL Mol. Cell. Biol. 8:5541-5544(1988).
 CC -1- FUNCTION: RECEPTOR FOR ACIDIC AND BASIC FIBROBLAST GROWTH FACTORS.
 CC POSSESSES A HIGHER AFFINITY FOR ACIDIC THAN FOR BASIC FGF'S.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN
 CC TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.cn/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC
 DR EMBL; M86441; AAA37286.1; -
 DR EMBL; M63503; AAA39377.1; -
 DR EMBL; M23362; AAA37285.1; -
 DR PIR; A31378; TYMSBK.
 DR PIR; A38429; A38429.
 DR HSSP; P11362; IFG1.
 DR MGD; MGI:95523; Fgfr2.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR001245; Tyr_Kin.
 DR Pfam; PF00047; Ig_3.
 DR Pfam; PF00069; Pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR SMART; SM00408; IgC2; 3.

RESULT 10

DR SMART; SM00219; TYRKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Glycoprotein; Tyrosine-protein kinase; ATP-binding;
 KW Transferase; Phosphorylation; Transmembrane; Signal;
 KW Immunoglobulin domain; Alternative splicing.
 FT SIGNAL 1 21
 FT CHAIN 22 821 FIBROBLAST GROWTH FACTOR RECEPTOR 2.
 FT DOMAIN 22 378 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 378 398 POTENTIAL.
 FT DOMAIN 398 821 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 55 114 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 172 238 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 271 349 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 481 770 PROTEIN KINASE.
 FT NP_BIND 487 495 ATP (BY SIMILARITY).
 FT BINDING 517 517 ATP (BY SIMILARITY).
 FT ACT_SITE 626 626 BY SIMILARITY.
 FT MOD_RES 657 657 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT DISULFID 62 107 POTENTIAL.
 FT DISULFID 179 231 POTENTIAL.
 FT DISULFID 278 342 POTENTIAL.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 123 123 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 147 147 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 265 265 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPIC 37 37 E -> G (IN SHORT ISOFORM).
 FT VARSPIC 38 152 MISSING (IN SHORT ISOFORM).
 FT VARSPIC 314 361 AAGVNTDKEIEVLRYNTFEDAGCYTCLAGNSIGISFHS
 FT VARSPIC 314 361 NYTGANOSAWLTVLPKQ (IN SHORT ISOFORM).
 FT CONFLICT 169 169 C -> Y (IN REF. 2).
 FT CONFLICT 187 187 S -> P (IN REF. 2).
 FT CONFLICT 187 187 C -> Y (IN REF. 2).
 SQ SEQUENCE 821 AA; 91983 MW; FCD828ADD61F4414 CRC64;

Query Match 4.3%; Score 102.5; DB 1; Length 821;
 Best Local Similarity 24.0%; Pred. No. 5.1; Indels 103; Gaps 18;
 Matches 78; Conservative 35; Mismatches 109;

DB 182 CFSNIFLASKL-----EAVEYAHISRLQADL--NCLSDL-----LKSSIQWRYVIN 225
 OY 491 CFGGVVAEAVGIDKDKPKKAVVYA-VKMLKDATEKDLSDLVSEMMKMGKHKNIIN 549
 OY 226 L--CGDDPLKSNFELVSELEKLGANMLETVKPPSKLEFRTYHHELRVREYVKLP 282
 DB 550 LIGACIDDDGL--YVIVEYASGNLREYLRARPPQME-----YSIDIRVPEEQWTF- 600
 OY 283 IRTNISLEAPPHNIQIFVGSAYFY-----LSQAFVKYIFNNSTIVQDFPANSKDTYSP 334
 DB 601 -----KDLVSCYQLARGMEYLAQKCIHRDLARVNLVTENNVMKIAQGLARDINNI 654
 OY 335 DEHEWATLLVPGIPGIGSISSACDVSDLOSKTRLVKMYNVEGFFYPSCITSHLSVCITY 394
 DB 655 DYIKATTNGRLP-----VKWMADEALFDRYVT--HSDVWSFG 690
 OY 395 -----AAELRWLIKGMFANKFDSKVDPLIKLAKLEEQORD-W 445
 DB 691 VLMWEIFTLGSGPYPGIPVEELFKLKEGH-----RMDKPTN-----CTNE-LVMWRDCW 740
 OY 436 IITLSEK-LF-----VDRNLTTS 453
 DB 741 HAVSPQRTFKQVLEDLRLITLT 765
 RESULT 11
 DPO3_UREPA

ID DPO3_UREPA STANDARD; PRT; 1442 AA.
 AC 09PGB4;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA POLYMERASE III POLC-TYPE (EC 2.7.7.7) (POLIII).
 GN POLC OR U0377.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OX NCBI_TaxID=134821;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Cassell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum";
 RL Nature 407:757-762(2000).
 CC -1- FUNCTION: REQUIRED FOR REPLICATIVE DNA SYNTHESIS. THIS DNA
 CC POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY (BY
 CC SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N-DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N-PYROPHOSPHATE + DNA(N).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. POLC
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC
 CC EMBL; AE002135; AAP30786.1; -
 CC InterPro; IPR000520; Exonuclease.
 CC InterPro; IPR003141; PNP_N.
 CC Pfam; PF00929; Exonuclease; 1.
 CC Pfam; PF02231; PNP_N; 1.
 DR SMART; SM00479; EXOIII; 1.
 DR SMART; SM00481; POLIII; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication; Hydrolyase;
 KW Nuclease; Exonuclease; Complete protease.
 FT DOMAIN 409 576 EXONUCLEASE.
 FT SEQUENCE 1442 AA; 166227 MW; 834C3EF722ACA6D1 CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 1442;
 Best Local Similarity 18.9%; Pred. No. 13; Indels 161; Gaps 27;
 Matches 100; Conservative 88; Mismatches 181;

DB 33 NVRRLEPQ-----KDIIVEYSLSTS-----PVRNRYTHVND-----VRYEVN 72
 OY 361 NVYQAVPEIAKSKKODLKI--YGLETEBELDQIPLVNL-----VQDMLDNATYVIFDIE 415
 DB 73 CGGIYE-----QEPLEIGKSELEIRRIIDLEDQDVYAMSDODIYITLNGY 119
 OY 416 TGLPFPNDEIEIFGAVIMQNNKGIQGIQFIKPIQOINEN---VTNLTNISQEMNN 471
 DB 120 A--QKLVSEKESFPLAYSLVHK-----DAIWERLHI 151
 OY 472 AIDKELALKIKEIFDDHILVHNGINPDININORLLKWKGLEPLKNSIDITLMISRAIN 531
 DB 152 AI-----YQHNITCIHYDR-KAPDTFYAMNKLAKCF--SNIF-IASKL 192
 OY 532 PKSHRLGAICKKYEYVDNDESAHRADYDAIVLADVFYKMKNNLENDGIIINLSINKL 591
 DB 193 EAVEYAHIS-----RLQADLNCISDLKSSIQWRYVINLGGDDPLKSNFELVSEL 243
 OY 592 OTTMLKNRBFQGWINILYKQANVAKMDYELVLSISHTDMYTR-----PTITTSPLANK 645


```

OY 244 KKLGNANLMTVKKPPNSKLERFTY---HHELRVY--PREVYKLPRTN-----ISKRAP 293
DB 646 DKLIIISNIH-----ESLINALSKNDEIKRLIQRYDITLTLSSQHLVYAKITTI 700
OY 294 ENIOFVGSAYFVLSQAFVKYIFNNISYVODFFANSSKDTY---SPDEHFMTLLIRVPPIG 350
DB 701 ENVQ-----KAFKRLIYALALEINKIITIIYSSSPYEFKDKKRYDVYVNTKGLEG 749
OY 351 EISSAOD--VSDIO---SKTRLKMNMYEGFFPSCGTGSLRVCVCIYGAALFEMLKDG 405
DB 750 KAHFANVYVPDLEIYDQKNALDELAYLE-----DEKLINLIINEN 791
OY 406 HMFANK-FDSKVPDILIKCLAEKLE---EQQRWITLPSKELFMDRLTJT 451
DB 792 PVHINSWEDSDIOPLKEGLYAPKKEGVQDKTIDYVHTAKIY-GENLPT 840

RESULT 12
YCF2_PINTH STANDARD; PRT; 2054 AA.
AC P41653;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL 244.6 KDA PROTEIN YCF2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024047; Pubmed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugita M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii."
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D17510; BAA04460.1; -
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA_1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 24460 MW; 6F5E92D078E33A9A CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2054;
Best Local Similarity 21.5%; Pred. No. 21;
Matches 72; Conservative 60; Mismatches 98; Indels 105; Gaps 20;

OY 15 KVFILFLTLW-LLSLKLNLNRLRFPQKDIYLVESLSTSPFYRNRVTHVKEDEYEVN 72
DB 864 QTFVLIYDLVRSFNLRLRLNP--FVREKRLYSIEISITPLPKEDQI-----VNEKN 914
OY 73 -CSGIYEOPELIGSLERRDITLEDVVMTSDCOYOTLRGYAKLVS-----K 126
DB 915 FCQPFKRSDBENNFQCFKR-----GFSSNGVLIQT-RSYQDDLSEKMSNK 962
OY 127 EKSFPILYSLVHKDAIVELRLHAINYCHNYICHIYDKRAPDTFVANNLAKCFSTI 186
DB 963 NEEIFPRT-----QDWFEVTECLKNKIYNE-----DIDGRS-----TISNSKEBQNI 1004

```

```

OY 187 FIKSKLEAFVYAHISRLQADLCLSLDKSSIQW-----KYVINLCQGDPLKSNFL 239
DB 1005 YRISQIDSI-----FSKW-----DLFTYMPWFTSAWCXYIENML-----LDT 1043
OY 240 VSELKRLKLAN-----MLETVPKPPNSKLERFTYHHELRVYVYKLPRTNISKRAPHN 295
DB 1044 LSLT-LLHGSNPFVSLQIWK-HNLLKR-NLWELSHPLMPDIOCKLRTNLI-N----- 1094
OY 296 IQFVGSATFVLSQAFVKYIFNNISYVODFFANSSKD 330
DB 1095 -----KEFFPSNFKDFPPYCKD 1112

RESULT 13
YCF2_TOBAC STANDARD; PRT; 2280 AA.
AC P09976; P09977;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL 267 KDA PROTEIN YCF2 (ORF 2280).
GN YCF2.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; easterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. BRIGHT YELLOW 4;
RA Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torazawa K., Meng B.Y., Sugita M.,
RA Denc H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
RT its gene organization and expression."
RL EMBO J. 5:2043-2049(1986).
RN [2]
RP REVISIONS.
RA Sugita M.;
RT Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z00044; CAA7427.1; -
DR EMBL: Z00044; CAA7438.1; -
DR PIR: A05204; A05204.
DR PIR: A05205; A05205.
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA_1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2280 AA; 266812 MW; E246D5F3D902C06D CRC64;

Query Match 4.2%; Score 101.5; DB 1; Length 2280;
Best Local Similarity 20.0%; Pred. No. 24;
Matches 88; Conservative 67; Mismatches 169; Indels 117; Gaps 20;

OY 9 KTLQGVILFLTLWLSLKLNLNRLRFPQKDIYLVESLSTSPFYRNRVTHVKEDEV 68
DB 761 KDTLNHRTIMKYTINQVLSLNRKSKQKWFEPILISRTFSRSMRDP-----DAYR 810

```


Job time: 205 sec

```

RESULT 15
VOO1_VACCC STANDARD; PRT; 666 AA.
AC P21093;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN O1.
GN O1.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; Pubmed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus."
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'."
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35027; AAA48053.1;
CC DR PIR: D42510; D42510.
SQ SEQUENCE 666 AA; 77577 MW; 46DF329DDEB66EF CRC64;

Query Match 4.2%; Score 99.5; DB 1; Length 666;
Best Local Similarity 22.0%; Pred. No. 6.4;
Matches 75; Conservative 50; Mismatches 111; Indels 105; Gaps 18;

QY 46 VEXSLSTSFVYRNRYTHVYKDEVRVYVNCSGIYEQEPLEIGKSLER----- 92
DB 292 VEVNISTYDE--RYRQFADEFRDYIM--IKERQITWOSGDRIRFRPRMSLRSTLIK 345
QY 93 RDIIIDLED-----DDVY-----AMTSPCDIYOTLRGYAQ 121
DB 346 KDTDSLEDILAHIDNARKKSKVISEDVERIISFRNLPCVYRRTKLSIDIDIKTI--MYL 403
QY 122 KIVSKKEKSFPTAYSLVYHKDAIMVERLIHAIYNOHNTYCIHYDRK-----ADTFKVA 175
DB 404 KIV-KDWSKSCALTLISAI--KGIWTDPTINTVLSK---ILHHRNVEFYVLTSVENKEIA 455
QY 176 MNNLAKCFNSNIFIAKSLKLEAVEYAHISRLOADLNC-----LSDLEKSSIQMKYVINLCG 228
DB 456 VNCGRSRLS-----LEFRELKSVRCDLRTDDGLDRLYDLTRALHGKINQNLIG 505
QY 228 QDF--PLKSNFELVSELKKNLNGANLLETVPKNSKLERFTYHHHELRRVPEYVXLPRTN 286
DB 506 QCMGQPLTEMLFENKKKKRLN--NLMEYIKIS---DMLVYGHSI-----EKTLPIDTS 554
QY 287 ISKEAPPHNIQIFVGSAYFVLSQAFVK-YIFNNSIVODEFA 326
DB 555 LSEFKLSVDTMS-----VLNDQYAKIVIFPNTIIEYIIA 587

```

Fri May 3 10:57:56 2002

us-09-645-192-2.rsp

Page 13

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:43:09 ; Search time 19.44 Seconds

(without alignments)
1775.055 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389

Sequence: 1 MKIFRCYKHTLQCKVILF.....DWITLPSEKLFMDENLTTS 453

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR_68:**
1: PIR1:**
2: PIR2:**
3: PIR3:**
4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863.5	36.1	428	2	A46293
2	711	29.8	400	2	A46297
3	326	13.6	445	2	T21262
4	323	13.5	425	2	T20207
5	315.5	13.2	454	2	T21261
6	296	12.4	401	2	T24929
7	294	12.3	422	2	T24188
8	275	11.5	402	2	T24930
9	245	10.3	459	2	T24742
10	229.5	9.6	454	2	T25382
11	220.5	9.2	489	2	T24013
12	219.5	9.2	470	2	T32137
13	219.5	9.2	486	2	T21566
14	203.5	8.5	472	2	T33384
15	202.5	8.5	753	2	T24745
16	197.5	8.3	448	2	D85042
17	193.5	8.1	478	2	T27714
18	188	7.9	467	2	T21408
19	187	7.8	447	2	T00906
20	182.5	7.6	513	2	T20068
21	178	7.5	367	2	T28892
22	175.5	7.3	406	2	F96571
23	168	7.0	395	2	B96735
24	166.5	7.0	434	2	T51450
25	132	5.5	630	2	T02524
26	131	5.5	384	2	T08940
27	126	5.3	371	2	T21819
28	116	4.9	895	2	E64431
29	108.5	4.5	661	2	S50734

30	108	4.5	702	2	B69498	hypothetical prote
31	107	4.5	717	2	S78177	hypothetical prote
32	107	4.5	1072	2	T37742	serine threonine-p
33	107	4.5	2265	2	T26183	hypothetical prote
34	106.5	4.5	1347	2	T41321	BTB domain and Ank
35	104	4.4	364	2	T48628	hypothetical prote
36	104	4.4	2485	1	H71621	serine/threonine-s
37	104	4.4	2628	2	S59413	probable membrane
38	103.5	4.3	588	2	H82877	hypothetical prote
39	102.5	4.3	415	2	T65223	heparin-binding fi
40	102.5	4.3	707	2	A54846	fibroblast growth
41	102.5	4.3	707	2	A58429	keratinocyte growt
42	102.5	4.3	769	2	S16236	fibroblast growth
43	102.5	4.3	820	2	S17295	fibroblast growth
44	102.5	4.3	821	1	TYMSBK	fibroblast growth
45	102.5	4.3	821	1	TYH022	fibroblast growth

ALIGNMENTS

RESULT 1
A46293
beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase
C:Species: Homo sapiens (man)
C:Date: 20-Oct-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A46293
R:Bierhuizen, M.F., Fukuda, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 9326-9330, 1992
A:Title: Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcN
A:Reference number: A46293; MIMD:93028457
A:Accession: A46293
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-428 <BI>
A:Cross-references: GB:M97347; NID:g183440; PID:AAA35919.1; PID:g183441
A:Experimental source: HU-60 cells
A>Note: sequence extracted from NCBI backbone (NCBI:115900, NCBI:115901)
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match	36.1%	Score 863.5	DB 2	Length 428
Best local similarity	43.1%	Pred. No. 6.2e-56		
Matches 187	Conservative	72	Mismatches 132	Indels 43
			Gaps 12	
QY	28	LKLLVRRLEF--POKDIIV- EYSLSTSPFYRNRTYKDE---VRY-----EV 71		
DB	1	MLRTLLRRRLFSYPTKYEMVLVLSLTFPSVLR--IHOKPEFVSRYHLELAGENPSSDI 57		
QY	72	NCSGIYOEPLFETIGK-----SLEIRRDIIIDEDDDVVAAMSDCDIYQTLRGVAQKIVS 125		
DB	58	NCRTKQVGDVNELOKVKLELTVKFKRP--KMPDPDIYNNMSDCSSFKRKRIYVEPLS 115		
QY	126	KEEKSPPIASLVYHKDAIWEVERLHAITYNHNHYICIHDKRAPDFTKYAAMNNLAKCFSN 185		
DB	116	KEEAEPPIASVYIYVHHKIMDLRLAIYMPNFCVHVDTSSEDSYLAAVGIAACSFN 175		
QY	186	IFPASTLEAVEVAHISRILOADLNCSDLSKSIOMKYVNLGQDPPLKSNEPELSEKK 245		
DB	176	VFAASRLAESVYASMSRVADLNCMDLYAMSAAMKYLINLGMDFPIETNEIYRKIL 235		
QY	246	LNCANMLEVTPKPNKSLREFTYHHELRVYEV--KLPFRINISKEAPENHIQIFVGA 303		
DB	236	LMGENMLETERFMSHKEEEM-----KRYEYVNGKL-TNNGIYVAMLPPLTPFSGSA 287		
QY	304	YFLSQAFYKIYINNSIYVDFEAFMSKDTSPDEPFAFLIYVPGIGETISRSNO-DVSD 362		
DB	288	YFVASEYVGYVQNEKIKIOLKMEWADDTSPDEIYLAITIQRIPEVPGSLPAHKKIDLSM 347		
QY	363	QSKTRVYKNNYEGEF-----YPSCTGSHLRSCYICGAELRMLIKDGMFANKEDSKYD 417		
DB	348	QAVARFVKMGYFEGDVSKGAPYRPGDVAVHVSVCIGAGADLMMMLKKNHLFANKFDVVD 407		

```
QY 418 PILIKLAETLEEQ 431
      | : | | | :
DB 408 LFAIQCLDEHLRHK 421
```

RESULT 2

A4029/
beta-[6-N-acetylpi]wosamivir]+transferrase (EC 2.4.1.-) = human

C:Species: Homo sapiens (man)

CiDate: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
CiAccession: A46297
R:Biernhuizen, M.F.; Mattei, M.G.; Fukuda, M.

Genes Dev. 7, 468-478, 1993
A:Title: Expression of the developmental I antigen by a cloned human cDNA encoding a me
A:Reference number: A46297; MUID:93194065
A:Accession: A46297

A/Status: preliminary
A/Molecule type: nucleic acid
A/Residues: 1-400 <BIE>

C: keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match: 29.8%; Score 711; DB 2; Length 400;
Best Local Similarity 43.7%; Pred. No. 9e-45;
Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps

	120/	CONSERVATIVE	120/	MISMATCHES	120/	INDELS	120/	GAPS
QY	108	SDCDIYQTLRGYAQKLVSKEEKSPFIANSVLVHKDAIMVERLIHAIYNOHNTYCIHDKR	167	108	SDCDIYQTLRGYAQKLVSKEEKSPFIANSVLVHKDAIMVERLIHAIYNOHNTYCIHDKR	167	108	SDCDIYQTLRGYAQKLVSKEEKSPFIANSVLVHKDAIMVERLIHAIYNOHNTYCIHDKR

DB 70 SSECRETQSHVITAPESKEEDDFLAYIMVHHHEDFRALFRALYMPQNYCVHDEK.129

QY 168 APDTEKVMANMLAKCFSNIFIAKLEAVEYAHISRLQADINCLSDILKSSIQWKYVINTLC 227

Query Match	13.6%;	Score 326;	DB 2;	Length 445;
Best Local Similarity	25.1%;	Pred. No. 2e-16;		
Matches 108;	Conservative 85;	Mismatches 176;	Indels 62;	Gaps 15;

```

25 LSLKLNLNRRLFPQKDIYLVESLSTSPFVRNRYTHVKCEV--RYEVNCSGIY----E 78
| | | | : | | | | | : | | : || | |
17 LVIPFLFYPLVPFR---ILSSINSSRRPOETSLISKIENDLLDDIDINCINFIENGSKN 73

```

```
79 QEPLEIGKSLERRDIDLEDDDVWAMTSDCDIYQTLRGYAQKLVSKEEKFPIAYSLV 130
    : | | | : : | | : : : | | ||| : | :
80 ENEFENHEDGCT      DEFTVWDDDDQGVZCETDEEWUWZCFADDFEECDFCEVCET 131
```

139 VHKD--AIMVER--LIHAIYNOHNICYIHYDRKAPDTFKYAMNNLAKCFSNFIASKLE 19

125 VYKELSOVTIQRYLFMSSIIYHPONEYCIAGVENSAPIFQNLKELSNCFNSIHIF-MKRP 18
194 AVEYAHISRLQADLNCISDLKSSIQWKYVINCQDFFPKSNFELYSELKTKIANGAMLE 25

184 PIDWGSHEIINSAYDCLEFSLHLSKSDWRVYQYLSGVDPILKTNLEMYQILKHLNG----- 23

254 TVKPPNSKLERFTYHHLELRVPVLEYVKKPIRTINISKEAPPHNIQIFVGSAYFVLSQAFVK 31

```

314 YIFNNSIVODFEAWSKDTYSPDEHFWATLI---RVGPISGETISRSAODYSDLOSCTRIV- 36
b
239 -----TANVEIK-PLYOQRL--RGKNETQSP--LPLEKSSLSLIPREAAAN 27

```

280 HLSSSSIPQLLEFL

```

330  ---KWNYY---EGFFPBCIGSHLSVCYIGAEELMWLIKQHWFAFKFDSKVDPLIK 42
340  PTDGWRYYISRDQIMSKPCHNYMKAGSCVFGIDVPRLLKSKALVAHFKYLLSEPEAYF 39

```

```

423 CLAEKEQR 433
    || 1:1:|
400 CL--LKERR 407

```

```

Db      5  IFGLFLLIFVTPLEPLVILVLEIFESQNPASIDIFS-----TSDCDIYQILRGYAOXLVSK 126
QY      75  GYEOEPLLEIGKSLERIRRDIDLEDDVYAM-----TSDCDIYQILRGYAOXLVSK 126
Db      45  KMPDGDVSLA-----RGALFEFDEDELLAQILKLSGQEMACAEOKIFGFEQPTSQ 98
QY      127  EKSFPPIAYSLVYHKDAIWEERLIHAIVNOHNYICIHDKRAPDFKVAANNNAKCFSN 186
Db      99  EELEFPIAYGMVHGDVQGLSLLSATYQPOQNOCLAVDGNSSVEFIQVRLMSRCYGN 158
QY      187  --FIASLEAVEYAHISRLQADNCLSDLKSSIQWKYVINLCGQDPKLSNFELVSEK 244
Db      159  QYFTTDEIRMGCEILTSV---FOCVDYLLAKLPSPDMKYYFOYLSGVADPAKLNSNEMIRILK 215
QY      245  KLGANMLLETVPKPPNSKLEERTYHHELRVRYEVYKLPJTNTISKEAPRNQIDFGSAY 304
Db      216  ALNG-----SERAEI--LPFEFYLRNKRKRPMSPLPLKYSL---SAT 253
QY      305  FVLSQAFVKYIFNNISIVQDFPAMSKDTYSPDEHFMTLLIRVPGI-----PGEISRSA 356
Db      254  F--SRKSANFMVNSEKYLEQIDFLRGTTCADESLMATIAGNPVKYIKFSELPMPGDPDKA 311
QY      357  QDVSDEL-QSKTRLYKM-----NYEGFFFP-----SCGSHLRSCVIGAELELR 400
Db      312  WIKHNYRIRIGKLGKYEENOKIDNGYVSRIOQYVNRAPVCKGYYIRLSGVGYDLPN 371
QY      401  LKDGHWFAKRPDSKYDPIILKCLAEKLEEQ 431
Db      372  LINRHELVAHKLVESYQPAFMCLVENSROK 402

RESULT  5
T21261
hypothetical protein F22D6.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21261
R:Milkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19397
A:Accession: T21261
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-454 <WILL>
A:Cross-references: EMBL:271262; PIDN:CA95816.1; GSPDB:GN00019; CESP:F22D6.12
A:Experimental source: clone F22D6
C:Genetics:
A:Gene: CESP:F22D6.12
A:Map position: 1
A:Introns: 62/3; 92/2; 190/3; 266/3; 291/3; 412/3

Query Match      13.2%; Score 315.5; DB 2; Length 454;
Best Local Similarity 24.3%; Pred. No. 1.2e-15;
Matches 111; Conservative 79; Mismatches 190; Indels 77; Gaps 16;

QY      8  FKHTLOOKVILFLTLMLSLKILNVRLEFPQKOIYLVYSTSPFYRNRYTHYKDEV 67
Db      6  FKSLTIAIFLFI-----YFSVESLFPKQ---EDKNVSKQFLKSTITTSDSY 52
QY      68  ---RYEVNGSIGYOEPLLEIGKSLERIRRDIDLEDDVYAMSDCDIYQTLRGYAOXLV 124
Db      53  LLDNMEINCSNII--KQYKTEKIDIMHDI---EEOLEFSCYNKQCLTKLTERFWTNDM 107
QY      125  SKEKSPPIAYSLVYHKDA-----IWEERLIHAIVNOHNYICJHYDR 166
Db      108  SAEKHEPLISGLVYKDLPOVPARMEFLKLFLEINLOVLFLLSSITHQNYCLAVGE 167
QY      167  KAPDTEKVAANNNAKCFSNFIASKLEAVEYAHISRLQADNCLSDLKSSIQWKYVINL 226
Db      168  NSAPIFONLREVSTCFSNVHF--MKRPISGSHHIDSYDCLLEFLSHLETWRYFOYL 226

```

```

QY      227  CGQDPKLSNFELVSELKLGANMLLETVPKPPNSKLEERTYHHELRVRYEVYKLPJRTN 286
Db      227  SGVDIPLKTNLEMLVOLKHLNSTWVELT---NYOARLITKNE----- 267
QY      287  ISKEAPRNHIQIFVSGAYFVLSQAFVKYIFNNISIVQDFPAMSKDTYSPDEHFMTLLIRV 346
Db      268  --NESP---LPLFKSSLIAIIPKRAANOLASSNTARKLLEFLMNEIDEGFWGTLPGNK 322
QY      347  ---GIPGEI--SRADVDLSQK---TRLVKNMY---EGFFPSCGSHLR--SVQYGA 396
Db      323  DQFNISGINSKDMWEYRONNNIFNPDPGWSYITSROQIDWPELCKYMKDDSCVFIG 382
QY      397  ELRWIKDGHWFANKRPSKYDPIILKCLAEKLEEQ 433
Db      383  DVPLRLTSKALVAHKKFYLKSEPEAYFCL---LKEHR 416

RESULT  6
T24929
hypothetical protein T15D6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24929
R:Dobson, R.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19356
A:Accession: T24929
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-401 <WILL>
A:Cross-references: EMBL:283125; PIDN:CAB05620.1; GSPDB:GN00019; CESP:T15D6.2
A:Experimental source: clone T15D6
C:Genetics:
A:Gene: CESP:T15D6.2
A:Map position: 1
A:Introns: 45/3; 75/2; 151/3; 227/3; 252/3; 296/2; 359/3

Query Match      12.4%; Score 296; DB 2; Length 401;
Best Local Similarity 23.4%; Pred. No. 2.8e-14;
Matches 86; Conservative 68; Mismatches 157; Indels 56; Gaps 7;

QY      61  THVKEVREYVNGSIGYOEPLLEIGKSLERIRRDIDLEDDVYAMTSCDIYQTLRGYA 120
Db      34  TOLLDDL--EINCTNII--QGLAKNEKOLEINTKMI---EDKLNSTDRCHTVKSMFRFN 86
QY      121  QKLVSKKEKSPPIAYSLVYHKDAIWEERLIHAIVNOHNYICIHDKRAPDFKVAANNNA 180
Db      87  EVPLSEEFARFPLSYGLVYKLSQVLFMLSSIYQPOQNEICIAVGENSASTFLILLELS 146
QY      181  KCFNSIFIAKLEAVEYAHISRLQADNCLSDLKSSIQWKYVINLCGQDPKLSNFELV 240
Db      147  DCFPN-----KRPITWGSYEIINSVYDCLKFLSKSNMKYFOYLSGVDIPLKTNLEMV 201
QY      241  SELKLGANMLLETVPKPPNSKLEERTYHHELRVRYEVYKLPJTNTISKEAPRNHIQIFV 300
Db      202  RIKKSLNGANVEIKYERRR-----LQONETESPPLPFK 237
QY      301  GSAYFVLSQAFVKYIFNNISIVQDFPAMSKDTYSPDEHFMTLLIRVPGI---IPGEISRAQ 357
Db      238  SSLSLIPKRAKANYLASSSIPOQLLEFLRNTVWADGGWGLFNGKGLFDVPGSLN---- 293
QY      358  DVSDLSQKTRLVKMYVYEGFFPSCGSHLRSCVIGAELELRWLKDGHWFAKRPDSKYD 417
Db      294  -----FEHQIMFESGCHNMKMGSCVFGIGDVSNNLQAKALVAHKKYLS 340
QY      418  PIIKCL 424
Db      341  PEAYFCL 347

RESULT  7
T22188

```


Db 73 KIEHFYSKYRRESEETAHVDCGRILSGDKDYLQTVSG--ENRIPMVENPNPDMSCSAVMD 130
 QY 133 -----IAYSLVYHDAIMVERLIIAIYNQNIYCIHIDRKAPDTFKYAMN 177
 Db 131 RIIPHDILRLPKGVAFARLVYDYDELVEQVOMSYHPQSFADIDKAPFRFKQMR 180
 QY 178 NLACFSNIFASRLAEVYAHISRLQADNCLSDLKSSIQWKYVNLGODFPPLKSNF 237
 Db 191 AMAACLPVLLLPQEPIDSGNHVNLALHNLCLALINKP-GWNYAALLQNHDLITFSY 249
 QY 238 ELVSELKLNANMLEVYKPPNSKLERFTYHEHLRVPYEVKLPITRNISKEAPRNIO 297
 Db 250 ELEQVTEMLGGANDVELL-PEAQRLDEENFKWD---PSLIKMFPPDESKYDETLLNEKIK 304
 QY 298 IFVGSAYFVLSQAFVKYIFNNSIVQDFP-AMSKDTYSPDEHFMATL--IRVPGIPGEISR 354
 Db 305 FSKGVOGSMRAAVDMWTRKVNLSYIDQMNQGRMGVDEMLSSLDLSAFLMGPHF-- 362
 QY 355 SAQVSDLOQSTRLVKNYIEGFYPPGCTGSHLR-SVCIYGAEMLRLIDGHWFKAFKD 413
 Db 363 TDQCLKEKEPEFMYQVWNEDESY---CASKYKRHNVCIIIGIELRVSASFPTLMFKML 419
 QY 414 SKVDPIILKCLAEKLEEO---ORDWITLPSEKLFMDRNLTTSS 453
 Db 420 PPFNDSITECTAELLNRTFNGQNDH---PLEEERY-KMYTVTS 459

RESULT 10
 T25382
 hypothetical protein T27F6.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T25382
 R: Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A: Reference number: Z20026
 A: Accession: T25382
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-454 <WTL>
 A: Cross-references: EMBL:Z82060; PDB:CAH04881.1; GSPDB:GN00019; CESP:T27F6.1
 A: Experimental source: clone T27F6
 A: Gene: CESP:T27F6.1
 A: Map position: 1
 A: Introns: 80/3; 191/1; 278/1; 325/3; 365/2; 407/3
 C: Superfamily: Caenorhabditis elegans hypothetical protein H41C03.3

Query Match 9.6%; Score 229.5; DB 2; Length 454;
 Best Local Similarity 22.7%; Pred. No. 2.5e-09;
 Matches 103; Conservative 68; Mismatches 202; Indels 81; Gaps 19;
 QY 17 FILFLTLMLSLKLLNVRLLFPQKDIYLVESLSTSPFYARNRYTHVKDEVREVNCSGI 76
 Db 10 FIL-----LVALICTLN-----WFLYNYQV-----YRNHWYSYTONSSFEQDIKAF 50
 QY 77 YEQEPLEIGKSLERRRDIIDLEDDV---VAAMTSDDIYQTLRG-AAOKLVSEKESFP 132
 Db 51 YPTSNKDV---FYRRRR---ETENYVNGCVLAGDIAIKTYVGEIRIKIAENESLIMS 103
 QY 133 -----IAYSLVYHDAIMVERLIIAIYNQNIYCIHIDRKAPDT 171
 Db 104 CEAVMDRLSRDHLRPLENGVAFARVYMDYELIEKHVEVSHYQPSFCFAIDKRAKE 163
 QY 172 FKVAMNNLAKGFSNIFI-----ASKLEAVEYAHISRLQADNCLSDLKSSIQWKY 222
 Db 164 FKRNQAAASCLPVLPLGRFFKNPIHDSVDSHGHTNL-AHYNCLRALINKP-GWNY 221
 QY 223 VINLGGODPPLKSNFELVSELKLNANMLETVKPPNSKLERFTYHEHLRVPYEVKLP 282
 Db 222 AILLQNHDLITRKYVELEKIFNWLGGANDV-AIRPELGRLDX-----KHFKWDPMSTK 273

QY 283 IRTNISKEAP---PHNIQIFVGSAYFVLSQAFVKYIFNNSIVQDFP-AMSKDTYSPDEHF 338
 Db 274 LERNSEIDPVLINTTLKRAKGVOSLSRAAVDMWTRFVDTLTITDQNNHGTIVDQF 333
 QY 339 WAT-LIRVPGIPGEIS-BSAODVSDLOQSTRLVKNYIEGFYPPGCTGSHLR-SVCIYG 394
 Db 334 TQAFQSDLELGMGPHHTDKCIKIGITTEGITRFAQTHD---QSKCASKRSRGICIMG 390
 QY 395 AALRLRLIDGHWFKAFKPSKVDPIILKCLAEKL 428
 Db 391 IEHLKMAKSEHLMENKVLPLFDYSITECTAELL 424

RESULT 11
 T24013
 hypothetical protein R07B7.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24013
 R: Harris, B.
 submitted to the EMBL Data Library, July 1996
 A: Reference number: Z19830
 A: Accession: T24013
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-489 <WIL>
 A: Cross-references: EMBL:Z75955; PDB:CAH00115.1; GSPDB:GN00023; CESP:R07B7.6
 A: Experimental source: clone R07B7
 A: Gene: CESP:R07B7.6
 A: Map position: 5
 A: Introns: 42/3; 80/2; 235/3; 287/1; 372/2; 416/1

Query Match 9.2%; Score 220.5; DB 2; Length 489;
 Best Local Similarity 22.0%; Pred. No. 1.3e-08;
 Matches 102; Conservative 77; Mismatches 209; Indels 75; Gaps 19;
 QY 22 TLWLISLKLNLNVRLLFP---QKDIYLVESL---STSPFYARNRYTH-----LEDDVYVAMTSDC 62
 Db 16 TLTKLEMSATRYKRLTDIDLADINTARVSLQDNGKNSPIILRTSQEKPRKRNTKKMD 75
 QY 63 ---VKDEVREYVNGSGIYEQEPLEIGKSLERRRDIID-----LEDDVYVAMTSDC 110
 Db 76 NYTTADESQF-LNCSEMIK-----NNKDIYELVYNNGRKLDNERLFELPMDC 122
 QY 111 -DIYQTLRGYAOKLVSEKESFPFIAYSLVYHDAIMVERLIIAIYNQNIYCIHIDRKAP 169
 Db 123 PSIKNRIYG---DMPSPFLPKRPIAFVRTIYKYLEQELALISYHPDNYVCFVMDSKST 179
 QY 170 DTFKVMNNLAKGFSNIFASRLAEVYAHISRLQADNCLSDLKSSIQWKYVINLGG 229
 Db 180 DRIKESVRLMSQFTVNVVLGKESVLSNGGQDPHFQCLLTIDR--KMDHAILQNF 237
 QY 230 DPFKLNFELVSELKLNANMLETVKPPNSKLERFTYHEHLR--VPYEVKLPITRN 287
 Db 238 DLITKTPYQSDISESLNTSITM-----GRDHGFSYVNNKAKKTPAGMKLFIKIEGV 290
 QY 288 SKAPAPRNIOIEVGSAYFVLSQAFVYIT---NNSLVQDFPMSKDTYSPDEHFMATL 343
 Db 291 PNEILNRNLIVKRSINEVYKVEYKSMEEKLMDIILIFD-DNDYVGVDEMLVQTLVE 349
 QY 344 RVGIGEI-SRAQVSDLOQSTRLVKNY-EGFFYPSCTGSHLR-SVCIYGAELRLW 400
 Db 350 NYLGLGQESNCTRNHNDI--LTRITMDFGSPNFDKESKRRHICIMGEYINME 407
 QY 401 LKDGHEMFANKEDSKVDPIILKCLAEKL-----EEQORDWIT 437
 Db 408 LIKSQGVIANKVMATPDEFCTIACRMIRKMTAGETPRNTQWLT 450

RESULT 12
 T32137

QY 73 -----CSGIYQEP-----LEIGKSLAIRRDIIDLEDDVAVAMSDC 110
 Db 76 AIISELKQDVFYKMLKTEFAIRYPRRPETHIDCGRLADKYLQTVSGKDRKIYENK 135
 QY 111 DIYTLNGVAKOKVSKSEKSP-----IAYSLVHKDAIWERLIIHAIYNOHNTIYHYDR 166
 Db 136 NLNNSCAIMNRILPSPGNDLKLKNGVAFARIYTDYEMIEKQVMQSHQNSCFEADIK 195
 QY 167 KAPDTFVAMNNLAKCSNIFIAKSLAEVAEHAHSRLQADLNCISDLKSSIQMKYVNL 226
 Db 196 KAPQFHERLMAACLPNVLPLPDESVDSAGHINILAHYINCLRVINKP-GNNVATLL 254
 QY 227 CGODEPLKSNELVSELKINGANMLETVPKPPNSKLEPFYHHELRVPYEVKLPF-- 283
 Db 255 QNHDAIVKSYEIQIYDWLGANDIE-ITPEAGRVDN-----KKFKMDPVSLKMFEN 306
 QY 284 RTINSKAPPHNIQIFVGSALFVLSQAFVKYIFNNSIVQDF-AWSKDTYSPDEHFWATL 342
 Db 307 ETGIDKNVLTSMKFAAGVQSLRGAVDMVMTVDLSTYINQWMESSGVDEQFIQS- 365
 QY 343 IRVP---GIPGEIS-----RSAQVSDLSQSKTRLVKMNY---YEGGFYP--SCTGSHL 387
 Db 366 FQVADLGMFGHFTDECLKOKRNTDYS-----RMSQMMYGSANHSFDIYVETABELL 419
 QY 388 RSVCIYGAEE 397
 Db 420 YNRTFLQVD 429

RESULT 15

T24745
 hypothetical protein T09E11.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24745
 R:McLay, K.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19930
 A:Accession: T24745
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-753 <NILL>
 A:Cross-references: EMBL:Z81147; PDB:CA03539.1; GSPDB:GN00019; CESP:T09E11.6
 A:Experimental source: clone T09E11
 C:Genetics:
 A:Gene: CESP:T09E11.6
 A:Map position: 1
 A:Introns: 34/2; 54/2; 191/3; 275/3; 315/3; 359/3; 380/3; 403/3; 426/3; 530/1; 617/1; 66

Query Match 8.5%; Score 202.5; DB 2; Length 753;
 Best Local Similarity 21.8%; Pred. No. 4.9e-07;
 Matches 96; Conservative 76; Mismatches 207; Indels 61; Gaps 16;

QY 21 LTLWL-----SLKLINVRLEFPQNDY--LVEYSLS-----TSPFVNRRTYHWKDEV 67
 Db 335 LNRWLLILIAASFISYFLVILLEKEBPYQTLSESTLSNCEGPTKODLHRHTDIADY 394
 QY 68 R---YEVNCSGIYQEP-----LEIGKSLAIRRDIIDLEDDVAVAMSDCDIYQTLRGYA 120
 Db 395 RERGFADKRGSYRRRPETAHVDCGRILAGDKPYLQSLTGTNRKTYENCNLNMSCAIR 454
 QY 121 QKLVSKEEK-----SPFIAYSLVHKDAIWERLIIHAIYNOHNTIYHYDRKAPDTEKVA 175
 Db 455 SRLPSNDNLRLPLKGIARFYKDYEFIEKQVQVSHPQNAFCVIDINASEEKRR 514
 QY 176 MNNIAKCFSNIFTASKLEAVYEAHISRLQADLNCISDLKSSIQMKYVNLGCGDPPKLS 235
 Db 515 MRALAAQMPVIYLADEDPVYSSGHNVLVNNCKLALDIP-GWNYALLQNHDLTKMS 573
 QY 236 NFELVSELKRLKNGANMLETVKPPNSKLEPFYHHELRVPYEVKLPF-----RTNI 287

Db 574 VYEMEQIFEMLGANDIEVT-----HEIGRVDYKLLKMDPMSLKLFINETEM 620
 QY 288 SKAAPPHNIQIFVGSALFVLSQAFVKYIFNNSIVQDF-AWSKDTYSPDEHFWATLIRVP 346
 Db 621 DKLLITTPMKIVKGWVHCSLSRASVEMFQKLDPSIFMHQLNGRGYDEQYFPILQANA 680
 QY 347 --GIPGEIS-RSAQVSDLSQSKTRLVKMNYEGGFYP--SC-TGSHLRVCIYGAELRW 400
 Db 681 EFGMPGHFTDECLQSGKTEFTFIRALW-----VPESKDTNMTTRAVCIIGLEHFGA 733
 QY 401 LIKDGHWFRANKFDSKYDPL 420
 Db 734 VASFTHLMFNKVSSSLDELI 753

Search completed: May 2, 2002, 07:45:39
 Job time: 150 sec

Fri May 3 10:57:56 2002

us-09-645-192-2.rpt

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 11:37:39 ; Search time 172.38 Seconds
(Without alignments)
6773.854 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362
Sequence: 1 atgaagattcaatcattta.....atctcactaccacatcatga 1362

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 10

Total number of hits satisfying chosen parameters: 456299

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_1101.*

```

1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1362	100.0	1362	22	AA500045 Human DNA encoding
2	349	25.6	781	22	AAH98678 Human EST-derived
3	183	13.4	186	21	AAI19209 Human secreted pro
4	25	1.8	31	22	AA500048 Human C2GNT3 PCR p
5	5	1.5	21	21	AA500046 Human C2GNT3 PCR p
6	20	1.5	32	22	AA500047 Human C2GNT3 sequ
7	20	1.5	1086	15	AA052335 Influenza type B H
8	19	1.4	364	21	AA31543 Plant microsatelli
9	19	1.4	1127	21	AA041094 Arabidopsis thalia
10	18	1.3	20	22	AA500049 Human C2GNT3 PCR p
11	18	1.3	255	22	AAH81918 Rat differential t

12	18	1.3	300	20	AA213168 Human gene express
13	18	1.3	528	21	AA11803 Aspergillus niger
14	18	1.3	592	20	AA12009 Neisseria meningit
15	18	1.3	592	21	AA81331 Neisseria meningit
16	18	1.3	807	20	AA12010 Complete ORF17 seq
17	18	1.3	807	20	AA12011 Complete ORF17 seq
18	18	1.3	807	20	AA12012 Complete ORF17 seq
19	18	1.3	1160	18	AA193285 Tomato mottle viru
20	18	1.3	1165	18	AA193285 Tomato mottle viru
21	18	1.3	1169	18	AA193285 Tomato mottle viru
22	18	1.3	1169	18	AA193285 Tomato mottle viru
23	18	1.3	1169	18	AA193285 Tomato mottle viru
24	18	1.3	1463	21	AA422240 Arabidopsis thalia
25	18	1.3	2456	21	AA481548 N. meningitidis pa
26	18	1.3	2602	18	AA193309 Tomato mottle viru
27	18	1.3	2843	21	AA257860 Protein regulating
28	18	1.3	3864	22	AA44664 Novel protein kina
29	18	1.3	4982	21	AA175062 Human ORF5 ORF617
30	18	1.3	5355	20	AA189230 Seq ID No: 28 of W
31	18	1.3	5355	20	AA189230 Seq ID No: 30 of W
32	18	1.3	5355	21	AA189230 Human TAD protein
33	18	1.3	5355	21	AA189230 Human glycosyl sul
34	18	1.3	7150	22	AA189230 Human T gene DNA.
35	18	1.3	81369	21	AA189230 Human gene signatu
36	18	1.3	240825	22	AA189230 Rice telomerase re
37	18	1.3	335913	22	AA189230 Human colon cancer
38	18	1.3	335913	22	AA189230 Human colon cancer
39	18	1.3	349980	21	AA189230 Probe #1182 used
40	18	1.3	349980	21	AA189230 B. burgdorferi ant
41	18	1.3	164976	19	AA189230 C glutamicum codin
42	18	1.2	294	22	AA189230 Human xliis gene fr
43	18	1.2	294	22	AA189230 B. burgdorferi ant
44	18	1.2	376	22	AA189230 Human secreted pro
45	18	1.2	555	22	AA189230 Streptococcus pneu
46	18	1.2	555	22	AA189230 S. epidermidis ope
47	18	1.2	724	22	AA189230 Arabidopsis thalia
48	18	1.2	732	22	AA189230 DNA encoding novel
49	18	1.2	816	20	AA189230 S. pneumoniae derl
50	18	1.2	837	20	AA189230 Arabidopsis thalia
51	18	1.2	917	21	AA189230 Xenopus alpha-sign
52	18	1.2	1096	18	AA189230 Treponema pallidum
53	18	1.2	1341	22	AA189230 Human cDNA sequenc
54	18	1.2	1417	22	AA189230 Enterococcus faeca
55	18	1.2	1453	22	AA189230 Corn cysteine1-TRN
56	18	1.2	1477	19	AA189230 Borrelia burgdorfe
57	18	1.2	1507	21	AA189230 Human polynucleoti
58	18	1.2	1769	18	AA189230 Human polynucleoti
59	18	1.2	1818	22	AA189230 Human cDNA sequenc
60	18	1.2	1863	22	AA189230 Human nerve growth
61	18	1.2	1923	22	AA189230 Homo sapiens fetal
62	18	1.2	2085	22	AA189230 Human cDNA clone B
63	18	1.2	2093	22	AA189230 Borrelia burgdorfe
64	18	1.2	2209	22	AA189230 S. epidermidis gen
65	18	1.2	2230	22	AA189230 Plasmid pBAC646.
66	18	1.2	2846	19	AA189230 Chicken genomic DN
67	18	1.2	2846	19	AA189230 4.757 kb fragment
68	18	1.2	2949	22	AA189230 Arabidopsis thalia
69	18	1.2	2991	22	AA189230 Arabidopsis ferula
70	18	1.2	3386	22	AA189230 Arabidopsis ferula
71	18	1.2	4009	16	AA189230 Human hpn3 cDNA se
72	18	1.2	4294	16	AA189230 Complete B-g gene
73	18	1.2	4757	16	AA189230 Chicken B-g gene
74	18	1.2	4757	22	AA189230 Polynucleotide seq
75	18	1.2	5156	18	AA189230 Enterococcus faecu
76	18	1.2	5156	18	AA189230 Streptococcus pneu
77	18	1.2	5156	21	AA189230 Polynucleotide seq
78	18	1.2	5874	22	AA189230 Human hpn3 cDNA se
79	18	1.2	7350	16	AA189230 Complete B-g gene
80	18	1.2	7350	16	AA189230 Chicken B-g gene
81	18	1.2	10928	20	AA189230 Polynucleotide seq
82	18	1.2	13574	20	AA189230 Enterococcus faecu
83	18	1.2	19390	19	AA189230 Streptococcus pneu
84	18	1.2	21170	20	AA189230 Polynucleotide seq

C 85	17	1.2	37940	20	AA201026	Partial mouse PGI
C 86	17	1.2	50000	22	AAF54867	Nucleotide sequence
C 87	17	1.2	53585	20	AA220251	Borrelia burgdorferi
C 88	17	1.2	80450	21	AAF22295	BAC containing rep
C 89	17	1.2	81145	22	AAF54868	Genomic nucleotide
C 90	17	1.2	130480	22	AAF25833	R. marinus bacteri
C 91	17	1.2	160552	22	AA020697	Human glycosyl sul
C 92	17	1.2	349980	22	AAH68532	C glutamicum codin
C 93	17	1.2	349980	22	AAH68533	C glutamicum codin
C 94	17	1.2	1644976	19	AAV21209	Methanococcus jann
C 95	16	1.2		20	AA083732	primer D8, to gene
C 96	16	1.2		21	AAH40578	SNP specific lower
C 97	16	1.2		22	AAH01917	biatrem resistance
C 98	16	1.2		22	AAH62884	Primer #3 for Full
C 99	16	1.2		24	AAV82610	Oligo-Amp used for
C 100	16	1.2		24	AAV82610	Plasmid pTR73-Pac
C 101	16	1.2		27	AAAT48159	Escherichia coli c
C 102	16	1.2		28	AAA37855	terramycin-like alph
C 103	16	1.2		30	AAQ11868	scat closing oligo
C 104	16	1.2		30	AAQ11871	Closing oligonucle
C 105	16	1.2		30	AAV32985	PCR primer NLC430
C 106	16	1.2		30	AAV23741	Cellulase variant
C 107	16	1.2		30	AAV27950	PCR primer for H.
C 108	16	1.2		31	AA500052	Human CGGnt3 sequ
C 109	16	1.2		34	AA606677	MLV variant rever
C 110	16	1.2		40	AA68464	Plasmid pBSstf1 c
C 111	16	1.2		40	AA68464	Circular plasmid e
C 112	16	1.2		55	AAK01785	Human cystic fibro
C 113	16	1.2		100	AAAT30896	Primer 26 for 95 K
C 114	16	1.2		111	AAAC14092	Human secreted pro
C 115	16	1.2		114	AAAC14952	Human secreted pro
C 116	16	1.2		121	AAI21554	Probe #11487 for g
C 117	16	1.2		121	AAI46845	Probe #15531 used t
C 118	16	1.2		121	AAI07248	Probe #7239 used t
C 119	16	1.2		167	AAV86342	EST clone GA327.
C 120	16	1.2		169	AAA41761	Human secreted exp
C 121	16	1.2		180	AAV89351	EST clone CT415.
C 122	16	1.2		190	AAI21532	Probe #11465 for g
C 123	16	1.2		190	AAI46824	Probe #15510 used
C 124	16	1.2		229	AAI07228	Probe #7219 used t
C 125	16	1.2		229	AAH50280	Bacterial 23S/5S R
C 126	16	1.2		229	AAH50335	Bacterial 23S/5S R
C 127	16	1.2		237	AAAC28394	Human secreted pro
C 128	16	1.2		279	AAI21667	Probe #11600 for g
C 129	16	1.2		279	AAI46956	Probe #15642 used
C 130	16	1.2		279	AAI07358	Probe #7349 used t
C 131	16	1.2		287	AAI23003	Probe #11936 for g
C 132	16	1.2		287	AAI47294	Probe #15980 used
C 133	16	1.2		287	AAI07699	Probe #7690 used t
C 134	16	1.2		288	AAI20039	Human gene signatu
C 135	16	1.2		291	AAI52671	Probe #21357 used
C 136	16	1.2		300	AAAC23456	Human secreted pro
C 137	16	1.2		300	AAAC00029	Human colon cancer
C 138	16	1.2		302	AAAX40308	Human secreted pro
C 139	16	1.2		307	AAAX43540	Human secreted exp
C 140	16	1.2		316	AAA44328	Human secreted exp
C 141	16	1.2		324	AAH93972	Human foetal cDNA,
C 142	16	1.2		339	AAAX41125	Human secreted pro
C 143	16	1.2		342	AAI63085	A diacylglycerol a
C 144	16	1.2		350	AAI63064	Human kidney relat
C 145	16	1.2		353	AAV88346	EST clone GA205.
C 146	16	1.2		381	AAAT67613	H. pylori cytoplas
C 147	16	1.2		385	AAAF64286	Novel human polyu
C 148	16	1.2		387	AAAF21955	Human breast and c
C 149	16	1.2		410	AAV20320	Probe (221) for m
C 150	16	1.2		427	AAAC79304	Human lung tumour-
C 151	16	1.2		437	AAH72101	Human cervical can
C 152	16	1.2		451	AAI12475	Probe #2408 for ge
C 153	16	1.2		451	AAI33830	Probe #2516 used t
C 154	16	1.2		451	AAI03386	Probe #3377 used t
C 155	16	1.2		455	AAI04414	Probe #4405 used t
C 156	16	1.2		457	AAAX21172	Poly nucleotide seq
C 157	16	1.2		461	AAI11944	Probe #1877 for ge
C 158	16	1.2		461	AAI33276	Probe #1962 used t
C 159	16	1.2		461	AAI01877	Probe #1868 used t
C 160	16	1.2		468	AAI10709	Probe #642 for gen
C 161	16	1.2		468	AAI14095	Probe #4028 for ge
C 162	16	1.2		468	AAI31967	Probe #653 used to
C 163	16	1.2		468	AAI35476	Probe #4162 used t
C 164	16	1.2		468	AAI006639	Probe #630 used to
C 165	16	1.2		468	AAI03947	Probe #3938 used t
C 166	16	1.2		469	AAAC00988	Human secreted pro
C 167	16	1.2		470	AAI27237	Probe #17070 for g
C 168	16	1.2		470	AAI53968	Probe #24674 used
C 169	16	1.2		478	AAI10185	Probe #118 for gen
C 170	16	1.2		478	AAI31434	Probe #120 used to
C 171	16	1.2		478	AAI00127	Probe #118 used to
C 172	16	1.2		478	AAE93487	CDNA encoding SRT
C 173	16	1.2		479	AAE93487	Human cervical can
C 174	16	1.2		480	AAI12337	Probe #2270 for ge
C 175	16	1.2		480	AAI33653	Probe #2379 used t
C 176	16	1.2		480	AAI02281	Probe #2242 used t
C 177	16	1.2		489	AAH72422	Human cervical can
C 178	16	1.2		489	AAH72990	Human cervical can
C 179	16	1.2		492	AAH70868	Human cervical can
C 180	16	1.2		522	AAV26098	S. aureus phosphor
C 181	16	1.2		524	AAAC01027	Human secreted pro
C 182	16	1.2		525	AAV26097	S. aureus phosphor
C 183	16	1.2		536	AAV21136	Poly nucleotide seq
C 184	16	1.2		537	AAH83938	DNA encoding a sta
C 185	16	1.2		558	AAH29611	Drosophila melanog
C 186	16	1.2		559	AAI40919	Probe #9605 used t
C 187	16	1.2		560	AAZ41409	Human normal pancr
C 188	16	1.2		587	AAI17949	Probe #1982 for ge
C 189	16	1.2		587	AAI42937	Probe #11623 used
C 190	16	1.2		589	AAH09408	Human CDNA clone (
C 191	16	1.2		594	AAI42533	Probe #11219 used
C 192	16	1.2		598	AAI39589	Probe #8275 used t
C 193	16	1.2		600	AAI17582	Probe #7515 for ge
C 194	16	1.2		600	AAI42497	Probe #11183 used
C 195	16	1.2		603	AAV79565	PRSV field strain
C 196	16	1.2		603	AAV79565	PRSV field strain
C 197	16	1.2		603	AAV79557	PRSV field strain
C 198	16	1.2		603	AAV79557	PRSV field strain
C 199	16	1.2		605	AAV79568	PRSV field strain
C 200	16	1.2		605	AAV53473	DNA encoding a sta
C 201	16	1.2		618	AAA43080	Human secreted exp
C 202	16	1.2		618	AAH77317	Igg Fab-BPI fusion
C 203	16	1.2		654	AAV60631	Beta-urogastrone -
C 204	16	1.2		658	AAAC6583	Human BPI CDNA seq
C 205	16	1.2		661	AAH08242	Fusarium venenatum
C 206	16	1.2		664	AAH71030	Human cervical can
C 207	16	1.2		667	AAH06953	Human CDNA clone (
C 208	16	1.2		699	AAZ80206	Human colon cancer
C 209	16	1.2		702	AAH07852	Human CDNA clone (
C 210	16	1.2		720	AAV13962	Human recombinant
C 211	16	1.2		728	AAV30915	Human immunoglobul
C 212	16	1.2		734	AAH98602	Human EST-derived
C 213	16	1.2		735	AAH40601	Yeast 2,5-Diamino-
C 214	16	1.2		755	AAAT16056	Human gene express
C 215	16	1.2		768	AAAD05339	Human secreted pro
C 216	16	1.2		769	AAAO1709	Human colon cancer
C 217	16	1.2		770	AAZ97293	Human prostate can
C 218	16	1.2		771	AAZ97293	Human prostate can
C 219	16	1.2		776	AAH06486	Human CDNA clone (
C 220	16	1.2		776	AAH08643	Human CDNA clone (
C 221	16	1.2		779	AAV36564	Human pituitary-tu
C 222	16	1.2		780	AAZ97592	Human prostate can
C 223	16	1.2		789	AAI01041	E. coli mature TEM
C 224	16	1.2		792	AAAC90773	Escherichia coli m
C 225	16	1.2		792	AAV37842	Escherichia coli m
C 226	16	1.2		792	AAV17986	Escherichia coli m
C 227	16	1.2		792	AAV32377	RTM with a mammal
C 228	16	1.2		792	AAI33046	E. coli TEM-1 beta
C 229	16	1.2		795	AAV37839	E. coli RTM beta-
C 230	16	1.2		795	AAV17983	Escherichia coli m
					AAZ32374	Escherichia coli R

231	16	1.2	795	22	AAD13043	E. coli TEM-1 beta
232	16	1.2	795	22	AAD13045	E. coli TEM-1 beta
233	16	1.2	798	22	AAH03926	Human cDNA clone (
234	16	1.2	803	22	AAH06590	Human cDNA clone (
235	16	1.2	808	22	AAH01936	Klebsiella oxytoca
236	16	1.2	817	19	AAV40858	Beta-lactamase fra
237	16	1.2	819	22	AAH03799	Human cDNA clone (
238	16	1.2	819	19	AAV63196	Human cDNA clone (
239	16	1.2	826	21	AAAC44565	2ea mays DNA fragm
240	16	1.2	827	21	AAAC61643	Human hair specific
241	16	1.2	835	20	AAV51719	DNA encoding a hum
242	16	1.2	840	19	AAV17985	Escherichia coli m
243	16	1.2	843	17	AAV37841	E. coli RTM beta-
244	16	1.2	843	20	AAZ32376	RTM with beta-glo
245	16	1.2	843	22	AAH05690	Human cDNA clone (
246	16	1.2	858	17	AAV37840	E. coli RTM beta-
247	16	1.2	858	19	AAV17984	Escherichia coli m
248	16	1.2	858	20	AAZ32375	Escherichia coli R
249	16	1.2	858	22	AAD13044	E. coli TEM-1 beta
250	16	1.2	860	17	AAH06539	Human cDNA clone (
251	16	1.2	861	17	AAT28560	Bacterial antiobiot
252	16	1.2	861	17	AAT31627	Tem gene encodes b
253	16	1.2	861	22	AAH01934	Neisseria meningit
254	16	1.2	861	22	AAH01937	Escherichia coli n
255	16	1.2	861	22	AAH01938	Escherichia coli n
256	16	1.2	861	22	AAH01939	Escherichia coli n
257	16	1.2	861	22	AAH01940	Escherichia coli n
258	16	1.2	861	22	AAH02018	Klebsiella pneumonia
259	16	1.2	864	22	AAH00262	Bartonella hensela
260	16	1.2	869	22	AAH72931	Human cervical can
261	16	1.2	870	17	AAT07318	Partial formyl-CoA
262	16	1.2	870	20	AAV64282	O. formigenes form
263	16	1.2	877	21	AAA44356	Human secreted exp
264	16	1.2	900	19	AAV38860	Bactericidal perme
265	16	1.2	908	20	AAV38925	Human regulatory p
266	16	1.2	933	22	AAH31708	Human olfactory re
267	16	1.2	933	22	AAH31768	Human olfactory re
268	16	1.2	948	21	AAAC48497	Arabisdopsis thalia
269	16	1.2	965	19	AAV43035	Streptococcus pneu
270	16	1.2	975	21	AAAC4939	Arabisdopsis thalia
271	16	1.2	1000	22	AAV94412	N. meningitidis (s
272	16	1.2	1037	21	AAAC59721	Human secreted pro
273	16	1.2	1075	21	AAAC34728	Arabisdopsis thalia
274	16	1.2	1080	15	AAO62337	Influenza type B H
275	16	1.2	1080	15	AAO62338	Influenza type B H
276	16	1.2	1080	15	AAO62339	Influenza type B H
277	16	1.2	1080	15	AAO62340	Influenza type B H
278	16	1.2	1080	15	AAO62341	Influenza type B H
279	16	1.2	1082	21	AAAC39223	Arabisdopsis thalia
280	16	1.2	1086	15	AAO62334	Influenza type B H
281	16	1.2	1086	15	AAO62347	Influenza type B H
282	16	1.2	1106	14	AAO49225	Plasmid pDS56/RBSI
283	16	1.2	1106	14	AAO49226	Plasmid pDS56/RBSI
284	16	1.2	1106	14	AAO49227	Plasmid pDS56/RBSI
285	16	1.2	1113	21	AAE15961	Human prostate can
286	16	1.2	1125	21	AAAC48273	Arabisdopsis thalia
287	16	1.2	1130	21	AAAC40921	Arabisdopsis thalia
288	16	1.2	1144	22	AAH29580	Drosophila melanog
289	16	1.2	1163	18	AAT87973	PCAL module M10-ii
290	16	1.2	1167	21	AAAS3885	Arabisdopsis Est sh
291	16	1.2	1169	22	AAT59058	Human polynucleoti
292	16	1.2	1212	5	AAV40167	Sequence encoding
293	16	1.2	1237	21	AAAC40215	Arabisdopsis thalia
294	16	1.2	1238	21	AAAC4544	Arabisdopsis thalia
295	16	1.2	1256	22	AAH28938	C. albicans apoptos
296	16	1.2	1258	18	AAT91739	DNA gyrase subunit
297	16	1.2	1259	21	AAAC42410	Arabisdopsis thalia
298	16	1.2	1266	22	AAH53212	S. epidermidis ope
299	16	1.2	1273	14	AAO42410	Tripartite fusion
300	16	1.2	1289	18	AAT88300	Beta-lactamase-WCS
301	16	1.2	1291	21	AAAC40558	Arabisdopsis thalia
302	16	1.2	1305	21	AAAC51959	Arabisdopsis thalia
303	16	1.2	1334	22	AAH15385	Human cDNA sequenc
304	16	1.2	1366	22	AAH15377	Human cDNA sequenc
305	16	1.2	1368	11	AAO06226	Amb a II clone enc
306	16	1.2	1368	19	AAV37864	Allergen Amb a II
307	16	1.2	1368	19	AAV13561	Ragweed pollen Amb
308	16	1.2	1385	21	AAAC42253	Arabisdopsis thalia
309	16	1.2	1392	22	AAO05340	Human secreted pro
310	16	1.2	1395	14	AAO51243	Ragweed Pollen All
311	16	1.2	1395	15	AAT32618	Ragweed Amb all al
312	16	1.2	1403	21	AAAC3040	Arabisdopsis thalia
313	16	1.2	1425	16	AAT20727	Polynucleotide seq
314	16	1.2	1426	20	AAV20336	Borrelia burgdorfe
315	16	1.2	1433	21	AAAC41836	Arabisdopsis thalia
316	16	1.2	1443	21	AAAC40612	Arabisdopsis thalia
317	16	1.2	1464	21	AAAC41677	Arabisdopsis thalia
318	16	1.2	1465	22	AAI60352	Human polynucleoti
319	16	1.2	1469	22	AAH42032	Disease treatment
320	16	1.2	1470	22	AAV58566	Human polynucleoti
321	16	1.2	1473	20	AAV28825	Rat neuronal lmed
322	16	1.2	1488	21	AAAC37961	Arabisdopsis thalia
323	16	1.2	1496	20	AAZ28233	Human calcium bind
324	16	1.2	1542	17	AAT42305	Male-specific DNA
325	16	1.2	1546	22	AAAC4564	Nucleotide sequenc
326	16	1.2	1598	22	AAO05306	Human secreted pro
327	16	1.2	1602	21	AAV07664	Fusarium venenatum
328	16	1.2	1610	20	AAV13498	Enterococcus faeca
329	16	1.2	1616	21	AAAC50714	Arabisdopsis thalia
330	16	1.2	1618	21	AAAC38969	Arabisdopsis thalia
331	16	1.2	1653	19	AAV39861	Bactericidal perme
332	16	1.2	1662	22	AAH02237	Tetragonococcus ha
333	16	1.2	1666	22	AAH15408	Human cDNA sequenc
334	16	1.2	1692	21	AAV65305	Human secreted pro
335	16	1.2	1707	21	AAAC39785	Arabisdopsis thalia
336	16	1.2	1722	21	AAH16832	Human cDNA sequenc
337	16	1.2	1727	21	AAZ47340	DD3 (dopa decarbo
338	16	1.2	1729	21	AAAC36534	Arabisdopsis thalia
339	16	1.2	1768	20	AAV37405	Human secreted pro
340	16	1.2	1799	18	AAV59214	Influenza B/Panama
341	16	1.2	1799	20	AAV00775	Influenza virus B/
342	16	1.2	1799	22	AAO09588	Influenza virus B/
343	16	1.2	1802	18	AAV59218	Influenza B/Harbin
344	16	1.2	1802	22	AAV00779	Influenza virus B/
345	16	1.2	1802	22	AAO09592	Influenza virus B/
346	16	1.2	1810	13	AAO25593	Sequence encoding
347	16	1.2	1811	18	AAV59215	Influenza B/Nether
348	16	1.2	1811	20	AAV00776	Influenza virus B/
349	16	1.2	1811	22	AAO09589	Influenza virus B/
350	16	1.2	1813	10	AAV92766	CDNA encoding huma
351	16	1.2	1813	14	AAV52268	Human bactericidal
352	16	1.2	1813	14	AAV52268	Human bactericidal
353	16	1.2	1813	15	AAO67270	Recombinant bacter
354	16	1.2	1813	15	AAO67270	Recombinant bacter
355	16	1.2	1813	15	AAO67270	Recombinant bacter
356	16	1.2	1813	15	AAO67270	Recombinant bacter
357	16	1.2	1813	15	AAO67270	Recombinant bacter
358	16	1.2	1813	15	AAO67270	Recombinant bacter
359	16	1.2	1813	15	AAO67270	Recombinant bacter
360	16	1.2	1813	15	AAO67270	Recombinant bacter
361	16	1.2	1813	15	AAO67270	Recombinant bacter
362	16	1.2	1813	15	AAO67270	Recombinant bacter
363	16	1.2	1813	15	AAO67270	Recombinant bacter
364	16	1.2	1813	15	AAO67270	Recombinant bacter
365	16	1.2	1813	15	AAO67270	Recombinant bacter
366	16	1.2	1813	15	AAO67270	Recombinant bacter
367	16	1.2	1813	15	AAO67270	Recombinant bacter
368	16	1.2	1813	15	AAO67270	Recombinant bacter
369	16	1.2	1813	15	AAO67270	Recombinant bacter
370	16	1.2	1813	15	AAO67270	Recombinant bacter
371	16	1.2	1813	15	AAO67270	Recombinant bacter
372	16	1.2	1813	15	AAO67270	Recombinant bacter
373	16	1.2	1813	15	AAO67270	Recombinant bacter
374	16	1.2	1813	15	AAO67270	Recombinant bacter
375	16	1.2	1813	15	AAO67270	Recombinant bacter
376	16	1.2	1813	15	AAO67270	Recombinant bacter

C 377	16	1.2	1813	20	AA224943	Recombinant human	C 450	16	1.2	2557	18	AA768842	Photorehabus lunin
C 378	16	1.2	1813	20	AA211725	Human bactericidal	C 451	16	1.2	2573	19	AAV29929	Internal gene frag
C 379	16	1.2	1813	20	AA280157	Human bactericidal	C 452	16	1.2	2573	16	AAQ79138	Ras gene derived f
C 380	16	1.2	1813	20	AA118800	CDNA encoding huma	C 453	16	1.2	2577	14	AAQ36620	Expression plasmid
C 381	16	1.2	1813	20	AAAX05716	Human bactericidal	C 454	16	1.2	2583	14	AAQ46606	Plasmid pINT4id co
C 382	16	1.2	1813	21	AAAS44281	Bactericidal/perme	C 455	16	1.2	2681	17	AA130870	Engineered 95 kd p
C 383	16	1.2	1813	21	AAAF75800	DNA encoding a hum	C 456	16	1.2	2686	18	AA161831	Puc18. Synthetic.
C 384	16	1.2	1813	21	AAAS8978	DNA encoding a hum	C 457	16	1.2	2686	22	AAAF59053	Plasmid vector pUC
C 385	16	1.2	1813	21	AAAF62318	Human recombinant	C 458	16	1.2	2698	22	AAH28230	Nucleotide sequenc
C 386	16	1.2	1813	21	AAAF62832	Recombinant holopt	C 459	16	1.2	2700	20	AAH29680	C. elegans kinase
C 387	16	1.2	1813	21	AAAF13265	Human bactericidal	C 460	16	1.2	2704	20	AAZ06758	Vector pUC28 nucle
C 388	16	1.2	1813	21	AAAF57575	Human bactericidal	C 461	16	1.2	2728	18	AAH79833	Vector pUC28 modu
C 389	16	1.2	1813	21	AAZ444274	Human BPI DNA. Ho	C 462	16	1.2	2731	21	AAAF14802	PAO-3 modular ve
C 390	16	1.2	1813	22	AAAF00568	Human BPI DNA. Ho	C 463	16	1.2	2739	20	AAAF10677	Nucleotide sequenc
C 391	16	1.2	1813	22	AAAF01832	Human recombinant	C 464	16	1.2	2747	22	AAH15001	BslI methylase gen
C 392	16	1.2	1813	22	AAAF085062	Human BPI protein	C 465	16	1.2	2753	22	AAH72608	Human CDNA sequenc
C 393	16	1.2	1813	22	AAAF085141	Human BPI protein	C 466	16	1.2	2754	21	AAH29008	Human CDNA sequenc
C 394	16	1.2	1813	22	AAAF085142	Human BPI protein	C 467	16	1.2	2754	21	AAH29008	Human CDNA sequenc
C 395	16	1.2	1813	22	AAAF25329	DNA encoding a hum	C 468	16	1.2	2755	13	AAAF03056	Vector pNN03. Syn
C 396	16	1.2	1813	22	AAAF91090	Human BPI coding s	C 469	16	1.2	2838	22	AAAF11388	Vector pTF74-N1-h
C 397	16	1.2	1813	22	AAAF31893	Human BPI coding s	C 470	16	1.2	2856	22	AAAF13388	Expression vector
C 398	16	1.2	1813	22	AAAF23641	Human BPI coding s	C 471	16	1.2	2866	22	AAAF16312	Human CDNA sequenc
C 399	16	1.2	1813	22	AAAF23641	Human BPI coding s	C 472	16	1.2	2869	22	AAAF1387	Expression vector
C 400	16	1.2	1813	22	AAAF68721	Human BPI nucleoti	C 473	16	1.2	2870	22	AAAF3095	Nucleotide sequenc
C 401	16	1.2	1813	22	AAAF04753	rBPI CDNA. Uniden	C 474	16	1.2	2871	19	AAAF2404	PT218 Vector. Syn
C 402	16	1.2	1814	18	AAAF59217	CDNA encoding reco	C 475	16	1.2	2919	18	AAH51124	Human homeoprotein
C 403	16	1.2	1814	20	AAAF00778	Influenza A/Shanha	C 476	16	1.2	2927	17	AAAF29158	S. epidermidis gen
C 404	16	1.2	1814	21	AAAF65682	Influenza virus B/	C 477	16	1.2	2927	17	AAAF29158	Plasmid pPRP. Syn
C 405	16	1.2	1814	21	AAAF47118	Human BPI CDNA SEQ	C 478	16	1.2	2927	17	AAAF11789	Plasmid pPRP. Syn
C 406	16	1.2	1814	22	AAAF09591	Reporter gene cons	C 479	16	1.2	2939	16	AAAF87350	Plasmid BGINV. Sy
C 407	16	1.2	1824	21	AAAF71128	Influenza virus B/	C 480	16	1.2	2943	19	AAAF72923	Plasmid pL-mu-smc
C 408	16	1.2	1861	22	AAAF16788	Human ORF2 ORF2683	C 481	16	1.2	2946	21	AAAF0704	Spor-rel coding seq
C 409	16	1.2	1866	19	AAAF53502	Human CDNA sequenc	C 482	16	1.2	2946	21	AAAF5650	Nucleotide sequenc
C 410	16	1.2	1910	22	AAAF13549	DNA encoding a Sta	C 483	16	1.2	2958	20	AAAF22250	Nucleotide sequenc
C 411	16	1.2	1947	18	AAAF7960	Human CDNA sequenc	C 484	16	1.2	2958	20	AAAF22250	Plasmid pPRP. Syn
C 412	16	1.2	1956	14	AAAF01235	PMCS cloning vecto	C 485	16	1.2	2973	19	AAAF1648	Plasmid pPRP. Syn
C 413	16	1.2	1956	20	AAAF20630	Sequence encoding	C 486	16	1.2	2994	19	AAAF60086	CDNA encoding a hu
C 414	16	1.2	1985	14	AAAF051083	Polynucleotide seq	C 487	16	1.2	2994	20	AAAF62697	Human TRAF2-associ
C 415	16	1.2	2007	22	AAAF15514	H. somnus lipa gen	C 488	16	1.2	3003	11	AAAF005745	Plasmid pL-mu-smc
C 416	16	1.2	2009	10	AAAF91727	Human CDNA sequenc	C 489	16	1.2	3018	20	AAAF4626	Novel protein kina
C 417	16	1.2	2064	16	AAAF080535	ss Coding strand o	C 490	16	1.2	3018	20	AAAF39905	Plasmid pGEM (RTM)
C 418	16	1.2	2064	16	AAAF080535	Paralichthys oliva	C 491	16	1.2	3034	21	AAAF29224	Human cell signal
C 419	16	1.2	2148	14	AAAF46026	Paralichthys oliva	C 492	16	1.2	3049	15	AAAF070890	Tomato spotted wil
C 420	16	1.2	2148	14	AAAF46026	Transglutaminase (C 493	16	1.2	3066	21	AAAF1641	PGM47 containing
C 421	16	1.2	2148	16	AAAF46026	Transglutaminase (C 494	16	1.2	3072	22	AAAF16011	Human CDNA sequenc
C 422	16	1.2	2148	16	AAAF080536	Paralichthys oliva	C 495	16	1.2	3098	22	AAAF3096	Nucleotide sequenc
C 423	16	1.2	2163	18	AAAF95753	Paralichthys oliva	C 496	16	1.2	3104	10	AAAF80296	Vector pASK75 frag
C 424	16	1.2	2178	20	AAAF22591	Arabidopsis SCAREC	C 497	16	1.2	3122	18	AAAF12724	DNA target sequenc
C 425	16	1.2	2187	21	AAAF14722	Arabidopsis thalia	C 498	16	1.2	3122	19	AAAF9219	Plasmid pG2B. Sy
C 426	16	1.2	2234	22	AAAF81507	Nucleotide sequenc	C 499	16	1.2	3122	20	AAAF22063	Nucleotide sequenc
C 427	16	1.2	2284	20	AAAF17619	White lupine secre	C 500	16	1.2	3130	19	AAAF46334	Universal reporter
C 428	16	1.2	2284	20	AAAF22595	Arabidopsis thalia	C 501	16	1.2	3138	12	AAAF11712	Shuttle vector pMW
C 429	16	1.2	2286	22	AAAF18022	Human CDNA sequenc	C 502	16	1.2	3153	19	AAAF65523	Plasmid DNA encodi
C 430	16	1.2	2321	21	AAAF51029	Pre-selection open	C 503	16	1.2	3159	22	AAAF75591	Human anti-HER2/ne
C 431	16	1.2	2335	19	AAAF52408	Streptococcus pneu	C 504	16	1.2	3160	19	AAAF6480	Human anti-HER2/ne
C 432	16	1.2	2337	21	AAAF51030	Pre-selection open	C 505	16	1.2	3165	21	AAAF51642	PGM51 containing
C 433	16	1.2	2349	18	AAAF60796	ORF5 2-5 of PRRSV	C 506	16	1.2	3175	20	AAAF4648	Human CDNA sequenc
C 434	16	1.2	2352	18	AAAF60796	ORF5 2-5 of PRRSV	C 507	16	1.2	3175	20	AAAF22066	DNA encoding gluta
C 435	16	1.2	2365	15	AAAF62815	Riboflavin 7 gene	C 508	16	1.2	3188	11	AAAF06310	Sequence of plasmid
C 436	16	1.2	2365	15	AAAF62815	Nucleotide sequenc	C 509	16	1.2	3190	19	AAAF32980	Tn7 target plasmid
C 437	16	1.2	2422	12	AAAF01711	Shuttle vector pMW	C 510	16	1.2	3201	22	AAAF14635	Human CDNA sequenc
C 438	16	1.2	2427	12	AAAF00460	Plasmid pL01223 u	C 511	16	1.2	3213	16	AAAF9870	Human CDNA sequenc
C 439	16	1.2	2462	21	AAAF74638	Plasmid pSP72. Es	C 512	16	1.2	3213	16	AAAF9870	Lactobacillus curv
C 440	16	1.2	2463	22	AAAF58206	Human CDNA sequenc	C 513	16	1.2	3216	21	AAAF49993	Plasmid pGNI for e
C 441	16	1.2	2465	22	AAAF18231	Human CDNA sequenc	C 514	16	1.2	3216	22	AAAF63882	Vector pGNI DNA se
C 442	16	1.2	2469	21	AAAF62771	Human ORF5 ORF1826	C 515	16	1.2	3246	22	AAAF00463	Plasmid pL012226 u
C 443	16	1.2	2489	16	AAAF98688	Lactobacillus curv	C 516	16	1.2	3250	22	AAAF17844	Human CDNA sequenc
C 444	16	1.2	2491	22	AAAF60311	Human polynucleoti	C 517	16	1.2	3253	17	AAAF43439	DNA sequence of th
C 445	16	1.2	2532	12	AAAF011202	Plasmid pINT4id en	C 518	16	1.2	3266	17	AAAF11192	Plasmid pASK75 com
C 446	16	1.2	2537	22	AAAF58525	Human polynucleoti	C 519	16	1.2	3276	11	AAAF06311	Sequence of plasmid
C 447	16	1.2	2528	21	AAAF16147	Human prostate can	C 520	16	1.2	3327	20	AAAF86263	Plasmid pGHR1-297A
C 448	16	1.2	2528	22	AAAF14419	Human CDNA sequenc	C 521	16	1.2	3331	22	AAAF84434	pBC6-01, containi
C 449	16	1.2	2539	12	AAAF15024	Rat prolactin poly	C 522	16	1.2	3331	21	AAAF29245	Nucleotide sequenc

C 523	16	1.2	3332	21	AA066453	Human secreted pro	C 596	16	1.2	3737	19	AAV43796	Sequence ID No.1 f
C 524	16	1.2	3343	11	AA004655	Plasmid p10169 enc	C 597	16	1.2	3752	22	AAV55130	Nucleotide sequenc
C 525	16	1.2	3344	21	AA039495	Transgenic unc-119	C 598	16	1.2	3754	16	AA057880	Intermediate plasm
C 526	16	1.2	3363	20	AA0208715	Alpha horidothionin	C 599	16	1.2	3754	16	AA085420	Plasmid pREC1. S
C 527	16	1.2	3365	20	AA0208716	Alpha horidothionin	C 600	16	1.2	3759	15	AA057881	Intermediate plasm
C 528	16	1.2	3369	22	AA068255	PCRHR-4 445K const	C 601	16	1.2	3769	16	AA097492	Plasmid pREC2. N
C 529	16	1.2	3379	22	AA066992	Vector pCGMT. Syn	C 602	16	1.2	3784	11	AA053265	Sequence encoding
C 530	16	1.2	3383	11	AA006309	Sequence of plasmid	C 603	16	1.2	3785	15	AA073378	STB5-CRTR chimera
C 531	16	1.2	3383	21	AA051632	Plasmid pGM678 con	C 604	16	1.2	3789	15	AA044362	Sequence of bovine
C 532	16	1.2	3384	21	AA023087	CNA of vector pGF	C 605	16	1.2	3789	21	AA039782	Plasmid pBIMN nu
C 533	16	1.2	3400	15	AA071366	E.coli/S.cerevisia	C 606	16	1.2	3794	19	AAV00680	Plasmid pTRH4 con
C 534	16	1.2	3403	9	AA080956	Plasmid pDS5/RBSII	C 607	16	1.2	3796	21	AA027831	Vector plasmid pCM
C 535	16	1.2	3414	12	AA012785	pDS5/RBSII-2 sequ	C 608	16	1.2	3803	11	AA003155	Sequence of the ch
C 536	16	1.2	3414	14	AA034609	Expression plasmid	C 609	16	1.2	3819	18	AAV78825	Kappa light chain
C 537	16	1.2	3415	14	AA012784	pDS5/RBSII-1 sequ	C 610	16	1.2	3819	19	AAV39266	Plasmid pRC65 nuc
C 538	16	1.2	3415	14	AA034608	Expression plasmid	C 611	16	1.2	3819	20	AAV22020	Nucleotide sequenc
C 539	16	1.2	3416	9	AA080958	Plasmid pDS5/RBSI	C 612	16	1.2	3820	11	AA003517	Plasmid pIGHL enco
C 540	16	1.2	3416	11	AA005302	Sequence of plasmid	C 613	16	1.2	3822	19	AAV13171	Complete DNA sequ
C 541	16	1.2	3416	12	AA012783	pDS5/RBSII sequen	C 614	16	1.2	3832	15	AA062676	Plasmid pASK60-St
C 542	16	1.2	3416	14	AA034607	Expression plasmid	C 615	16	1.2	3833	10	AA091060	Sequence of plasmid
C 543	16	1.2	3418	20	AA0206440	Plasmid pR27075. f	C 616	16	1.2	3840	15	AA073376	STB5-CRTR chimera
C 544	16	1.2	3419	21	AA014902	Nucleotide sequenc	C 617	16	1.2	3840	15	AA073377	STB5-CRTR chimera
C 545	16	1.2	3420	22	AA020728	Expression vector	C 618	16	1.2	3850	9	AA016341	Sequence of new pl
C 546	16	1.2	3420	22	AA063718	Expression plasmid	C 619	16	1.2	3851	18	AAV79857	Yellowtail tuna DN
C 547	16	1.2	3427	11	AA005304	Sequence of plasmid	C 620	16	1.2	3851	18	AAV90019	Yellow tail/fin tu
C 548	16	1.2	3427	11	AA006313	Sequence of plasmid	C 621	16	1.2	3853	19	AAV40006	Plasmid pCTM. Chi
C 549	16	1.2	3438	19	AAV39080	S. viridochromogen	C 622	16	1.2	3854	15	AA067221	Plasmid pSEC-cyc/c
C 550	16	1.2	3440	11	AA006303	Sequence of plasmid	C 623	16	1.2	3858	22	AA004947	Plasmid pRK50 used
C 551	16	1.2	3440	12	AA013728	pDS5/RBSII. 6XHis	C 624	16	1.2	3861	15	AAV13174	Complete DNA sequ
C 552	16	1.2	3444	21	AA015630	Plasmid pGM679 con	C 625	16	1.2	3867	20	AAV20067	DNA encoding gluta
C 553	16	1.2	3450	22	AA066993	Vector pGM679-1b.	C 626	16	1.2	3871	22	AA010008	Plasmid pSFM 24 to
C 554	16	1.2	3464	22	AA091232	Human DHR gene ex	C 627	16	1.2	3876	21	AAV98037	Expression vector
C 555	16	1.2	3476	19	AAV48232	Vector plasmid yce	C 628	16	1.2	3876	21	AAV98038	Expression vector
C 556	16	1.2	3476	19	AAV33299	E.coli-Corynebacte	C 629	16	1.2	3878	14	AA040029	Sequence of plasmid
C 557	16	1.2	3480	21	AA098011	Expression vector	C 630	16	1.2	3878	16	AA076041	Retrovirus vector
C 558	16	1.2	3488	18	AA078782	pCALO-2 modular ve	C 631	16	1.2	3878	21	AA089876	Plasmid pSep-pgt-
C 559	16	1.2	3519	21	AA098035	Expression vector	C 632	16	1.2	3879	21	AA098015	Expression vector
C 560	16	1.2	3547	7	AA060846	Plasmid sequence e	C 633	16	1.2	3879	21	AA098016	Expression vector
C 561	16	1.2	3550	9	AA081540	Sequence of pTRX-2	C 634	16	1.2	3879	21	AA098017	Expression vector
C 562	16	1.2	3550	13	AA031623	pTRX-2 containing	C 635	16	1.2	3879	21	AA098021	Expression vector
C 563	16	1.2	3552	11	AA005875	pTRX-2 plasmid seq	C 636	16	1.2	3879	21	AA098036	Expression vector
C 564	16	1.2	3557	7	AA060801	Human pro-growth h	C 637	16	1.2	3880	19	AAV39242	Plasmid pCK7-96 nu
C 565	16	1.2	3558	21	AA075006	Nucleotide sequenc	C 638	16	1.2	3881	18	AAV78801	Kappa light chain
C 566	16	1.2	3585	18	AA090399	Plasmid pALV1-781	C 639	16	1.2	3881	20	AA021996	Nucleotide sequenc
C 567	16	1.2	3604	21	AA015724	Expression vector	C 640	16	1.2	3885	21	AA098013	Expression vector
C 568	16	1.2	3618	7	AA021890	Nucleotide sequenc	C 641	16	1.2	3886	21	AAV13175	Complete DNA sequ
C 569	16	1.2	3621	20	AA060847	Human pre-prolactin	C 642	16	1.2	3888	19	AAV15008	Nucleotide sequenc
C 570	16	1.2	3621	22	AA055129	Nucleotide sequenc	C 643	16	1.2	3897	21	AA098019	Expression vector
C 571	16	1.2	3623	14	AA041297	BMP-2 expression v	C 644	16	1.2	3900	21	AA098030	Expression vector
C 572	16	1.2	3628	7	AA060848	Plasmid sequence e	C 645	16	1.2	3903	22	AA076078	Feline IL-16 encod
C 573	16	1.2	3632	13	AA027425	pALTRX/EK/TL11-de	C 646	16	1.2	3907	20	AAV13280	Enterococcus faeca
C 574	16	1.2	3632	13	AA027476	pALTRX/EK/TL11-del	C 647	16	1.2	3908	22	AA013152	Escherichia coli T
C 575	16	1.2	3632	15	AA056898	E.coli thio-redoxin	C 648	16	1.2	3912	21	AA098018	Expression vector
C 576	16	1.2	3632	15	AA044670	Thio-redoxin-Interi	C 649	16	1.2	3912	21	AA098027	Expression vector
C 577	16	1.2	3632	16	AA093133	Vector for thio-red	C 650	16	1.2	3919	21	AA098028	Expression vector
C 578	16	1.2	3632	16	AA090770	Thio-redoxin/des-pr	C 651	16	1.2	3926	11	AA003736	Sequence of plasmid
C 579	16	1.2	3634	22	AA081730	Human protease and	C 652	16	1.2	3926	22	AA010016	Plasmid pSFM 9 to
C 580	16	1.2	3648	22	AA0606201	Pestivirus NS5A f	C 653	16	1.2	3927	21	AA098026	Expression vector
C 581	16	1.2	3664	15	AA062675	Plasmid pASK46 for	C 654	16	1.2	3927	21	AA098032	Expression vector
C 582	16	1.2	3681	12	AA013578	Plasmid pKSEUS. S	C 655	16	1.2	3930	22	AA098093	Mus musculus goose
C 583	16	1.2	3698	18	AA070518	Sequence of vector	C 656	16	1.2	3932	21	AA098024	Expression vector
C 584	16	1.2	3699	20	AA020638	Vector pGPE. Ratt	C 657	16	1.2	3936	20	AA098023	Expression vector
C 585	16	1.2	3699	13	AA023445	Plasmid pGPE for c	C 658	16	1.2	3938	20	AA061041	Plasmid pPFA6-Kanx
C 586	16	1.2	3699	14	AA044179	Heavy chain minilo	C 659	16	1.2	3944	13	AA029750	PCMV vector. Synt
C 587	16	1.2	3699	17	AA037237	Plasmid pBSGFP exp	C 660	16	1.2	3945	21	AA098033	Expression vector
C 588	16	1.2	3699	18	AAV14340	Plasmid pGPE. Syn	C 661	16	1.2	3954	22	AA098034	Plasmid pSR-97 nuc
C 589	16	1.2	3699	18	AAV12534	Plasmid pGPE #2.	C 662	16	1.2	3954	21	AA098037	Nucleotide sequenc
C 590	16	1.2	3699	18	AAV02619	Unique BglII site	C 663	16	1.2	3954	22	AA062650	Plasmid pGHR1-29TW
C 591	16	1.2	3705	22	AA083391	Plasmid pGHR1. N	C 664	16	1.2	3955	19	AAV13173	Complete DNA sequ
C 592	16	1.2	3712	16	AA097490	Plasmid pCMH142 co	C 665	16	1.2	3956	21	AA064258	Plasmid pPR13/14 D
C 593	16	1.2	3723	22	AA010000	Retroviral control	C 666	16	1.2	3957	22	AA098034	Expression vector
C 594	16	1.2	3728	22	AA004929	Rat prolactin gene	C 667	16	1.2	3968	21	AA099981	pHSF70-1MG5 const
C 595	16	1.2	3735	12	AA014452		C 668	16	1.2	3976	13	AA025119	Expression vector

C 669	16	1.2	3976	22	AA086256	pgHRH1-44WTSK685 C	742	16	1.2	4173	19	AAV32663	Plasmid pGFP-TT re
C 670	16	1.2	3977	14	AA049223	Plasmid pDS56/RBS1	743	16	1.2	4176	10	AAV90709	Sequence of plasmid
C 671	16	1.2	3977	15	AA087695	Plasmid pDS56/RBS1	744	16	1.2	4184	21	AA16646	Human secreted pro
C 672	16	1.2	3977	15	AA087696	Plasmid pDS56/RBS1	745	16	1.2	4189	11	AA005397	Secretion Vector p
C 673	16	1.2	3977	15	AA087700	Plasmid pDS56/RBS1	746	16	1.2	4189	11	AA005397	Plasmid pGM712 con
C 674	16	1.2	3977	15	AA087701	Plasmid pDS56/RBS1	747	16	1.2	4193	21	AA051634	H. ghiliani1/B. me
C 675	16	1.2	3977	15	AA087702	Plasmid pDS56/RBS1	748	16	1.2	4201	21	AA097949	Human cancer assoc
C 676	16	1.2	3977	15	AA087703	Plasmid pDS56/RBS1	749	16	1.2	4201	21	AA078084	Human colon cancer
C 677	16	1.2	3977	15	AA087704	Plasmid pDS56/RBS1	750	16	1.2	4201	22	AAH33277	Feline IL-12p35 en
C 678	16	1.2	3977	15	AA087705	Plasmid pDS56/RBS1	751	16	1.2	4207	10	AAV92031	Sequence of recomb
C 679	16	1.2	3977	15	AA087706	Plasmid pDS56/RBS1	752	16	1.2	4214	21	AA062682	Vector pEP1ink6 D
C 680	16	1.2	3977	15	AA087707	Plasmid pDS56/RBS1	753	16	1.2	4214	21	AA059280	Nucleotide sequenc
C 681	16	1.2	3977	15	AA087708	Plasmid pDS56/RBS1	754	16	1.2	4222	10	AAV92038	Sequence of recomb
C 682	16	1.2	3977	15	AA087709	Plasmid pDS56/RBS1	755	16	1.2	4224	22	AAV55132	Nucleotide sequenc
C 683	16	1.2	3977	15	AA087710	Plasmid pDS56/RBS1	756	16	1.2	4229	22	AAH25896	Genetic informatio
C 684	16	1.2	3977	15	AA087711	Plasmid pDS56/RBS1	757	16	1.2	4249	19	AAV63466	Plasmid pCTME. C
C 685	16	1.2	3977	15	AA087712	Plasmid pDS56/RBS1	758	16	1.2	4257	20	AA022251	Nucleotide sequenc
C 686	16	1.2	3977	15	AA087686	Plasmid pDS56/RBS1	759	16	1.2	4259	14	AAV92002	pCR5-EGFP report
C 687	16	1.2	3977	15	AA087687	Plasmid pDS56/RBS1	760	16	1.2	4274	21	AA054144	Sequence of plasmid
C 688	16	1.2	3977	15	AA087688	Plasmid pDS56/RBS1	761	16	1.2	4276	12	AA006790	Plasmid pMW11. C
C 689	16	1.2	3977	15	AA087689	Plasmid pDS56/RBS1	762	16	1.2	4277	14	AA040280	Sequence of clone
C 690	16	1.2	3977	15	AA087689	Plasmid pDS56/RBS1	763	16	1.2	4277	21	AA089874	PCMV-IT nucleic ac
C 691	16	1.2	3977	15	AA087690	Plasmid pDS56/RBS1	764	16	1.2	4282	22	AAV83668	DNA encoding hgh.
C 692	16	1.2	3977	15	AA087684	Plasmid pDS56/RBS1	765	16	1.2	4283	18	AAV86449	Epidermal growth f
C 693	16	1.2	3977	15	AA087685	Plasmid pDS56/RBS1	766	16	1.2	4283	19	AAV86410	Plasmid pNRG1630 n
C 694	16	1.2	3977	15	AA073431	Plasmid pDS56/RBS1	767	16	1.2	4283	21	AA063227	Human ORF135
C 695	16	1.2	3977	15	AA087697	Plasmid pDS56/RBS1	768	16	1.2	4285	21	AA075780	Vector MS23-DRT10
C 696	16	1.2	3977	15	AA087691	Plasmid pDS56/RBS1	769	16	1.2	4299	21	AA027980	Vector ORF3-const
C 697	16	1.2	3977	15	AA087692	Plasmid pDS56/RBS1	770	16	1.2	4300	21	AA029135	pCR5-EGFP const
C 698	16	1.2	3977	15	AA087692	Plasmid pDS56/RBS1	771	16	1.2	4308	22	AA026449	DN05 Plasmid for
C 699	16	1.2	3977	15	AA087693	Plasmid pDS56/RBS1	772	16	1.2	4308	22	AAV26524	Plasmid pPUP2 enco
C 700	16	1.2	3977	15	AA087699	Plasmid pDS56/RBS1	773	16	1.2	4331	15	AA062391	Vector pVNC1. Syn
C 701	16	1.2	3977	15	AA087694	Plasmid pDS56/RBS1	774	16	1.2	4343	21	AAV75007	Nucleotide sequenc
C 702	16	1.2	3979	11	AA006512	Sequence of plasmid	775	16	1.2	4344	21	AAV35848	Vector pTGF67. S
C 703	16	1.2	3979	21	AA051643	pCMV00 with udp an	776	16	1.2	4357	22	AAV31389	Expression vector
C 704	16	1.2	3983	22	AA010265	pPC/AC vector DNA.	777	16	1.2	4366	16	AA085424	Plasmid pTREC3-P28
C 705	16	1.2	3984	16	AA070310	Vector plasmid PTO	778	16	1.2	4368	21	AA098020	Expression vector
C 706	16	1.2	3988	22	AAV87639	Human porphobilino	779	16	1.2	4369	21	AA095578	Plasmid pTREC1-P28
C 707	16	1.2	3988	22	AAV72857	pgm DNA. Homo sa	780	16	1.2	4377	16	AA085421	Expression vector
C 708	16	1.2	3989	21	AAV98025	Expression vector	781	16	1.2	4380	21	AAV98022	Nucleotide sequenc
C 709	16	1.2	3991	10	AAV91061	Sequence of plasmid	782	16	1.2	4382	22	AAV55131	Plasmid pBPM 5 to
C 710	16	1.2	3992	21	AAV97948	B. megaterium gluc	783	16	1.2	4386	22	AA010014	CRS-P30-CRS fusion
C 711	16	1.2	3999	20	AAV02664	Ep-892047 Seq ID 4	784	16	1.2	4390	21	AA055632	atb cloning vecto
C 712	16	1.2	4003	11	AAV06305	Sequence of plasmid	785	16	1.2	4396	21	AA055632	Eraduoxina-TRIP f
C 713	16	1.2	4004	11	AAV06306	Sequence of plasmid	786	16	1.2	4403	13	AA048899	DNA sequence of he
C 714	16	1.2	4005	14	AAV040418	Complete DNA sequ	787	16	1.2	4421	17	AAV58319	Expression vector
C 715	16	1.2	4009	19	AAV13172	Plasmid pTRH43 con	788	16	1.2	4432	15	AA074206	Expression vector
C 716	16	1.2	4009	19	AAV00679	Ep-892047 Seq ID 3	789	16	1.2	4432	16	AAV01113	Vaccine vector VJ
C 717	16	1.2	4019	20	AAV02663	Plasmid pCTM1. Ch	790	16	1.2	4432	22	AAV09426	Arabidopsis thalia
C 718	16	1.2	4026	19	AAV40007	Plasmid pMTM1. SY	791	16	1.2	4432	22	AAV09612	Murine anti-botuli
C 719	16	1.2	4032	19	AAV23237	Vector pOE60-Mact.	792	16	1.2	4433	18	AAV95023	Encodes amplicillin
C 720	16	1.2	4045	15	AAV070942	Plasmid pND311 con	793	16	1.2	4435	20	AAV86663	Plasmid pAN441 for
C 721	16	1.2	4045	22	AAV91149	Plasmid pOE60-Mact.	794	16	1.2	4443	8	AAV70627	Human cyclin DI-hu
C 722	16	1.2	4054	20	AAV01001	Eukaryotic express	795	16	1.2	4450	22	AAV50237	Human cyclin DI-hu
C 723	16	1.2	4059	16	AAV08475	Vector pAPEX-1. S	796	16	1.2	4451	21	AAV25637	Human cyclin DI-hu
C 724	16	1.2	4059	17	AAV41899	Apex-1 eukaryotic	797	16	1.2	4453	18	AAV69890	Human cyclin DI/hu
C 725	16	1.2	4059	17	AAV12239	Apex-1 eukaryotic	798	16	1.2	4454	19	AAV54118	Human cyclin DI/hu
C 726	16	1.2	4059	20	AAV19812	Vector pAPEX-1. S	799	16	1.2	4454	14	AAV33115	Plasmid pSV16B5, f
C 727	16	1.2	4059	20	AAV07338	Plasmid pDS7/RBS1	800	16	1.2	4454	22	AAV66462	Plasmid pCM1.8 co
C 728	16	1.2	4068	9	AAV080957	Sequence of plasmid	801	16	1.2	4454	18	AAV66462	Vector pGX104 DNA
C 729	16	1.2	4068	16	AAV90730	Plasmid pDS7/RBS1	802	16	1.2	4464	22	AAV66462	Plasmid pGX104 DNA
C 730	16	1.2	4074	21	AAV248267	Sequence of plasmid	803	16	1.2	4466	22	AAV66462	Plasmid pGX104 DNA
C 731	16	1.2	4088	19	AAV64255	Plasmid pTRNMT1 CO	804	16	1.2	4466	10	AAV80286	Plasmid pPR104-4.
C 732	16	1.2	4100	21	AAV50629	Plasmid pPK5/6 DNA	805	16	1.2	4466	15	AAV808637	pPR104-4 sequenc
C 733	16	1.2	4102	19	AAV64257	Recombinant plasmid	806	16	1.2	4477	13	AAV022962	Sequence of pHCV-3
C 734	16	1.2	4114	16	AAV090777	Plasmid pPK9/10 DN	807	16	1.2	4481	13	AAV021677	Sequence of pHCV-3
C 735	16	1.2	4118	18	AAV69188	Thioredoxin/IL-11	808	16	1.2	4481	14	AAV038235	HCV CRS-Core fusio
C 736	16	1.2	4118	18	AAV69189	Construct pGEM-HTR	809	16	1.2	4481	14	AAV038235	HCV CRS-Core expre
C 737	16	1.2	4145	14	AAV040279	Sequence of clone	810	16	1.2	4481	22	AAV32218	HCV CRS-Core expre
C 738	16	1.2	4145	14	AAV08673	Plasmid pS2gpt-S4.	811	16	1.2	4482	15	AAV022662	pUO200 expression
C 739	16	1.2	4163	13	AAV032349	Template clone Fab	812	16	1.2	4484	22	AAV63719	Expression plasmid
C 740	16	1.2	4163	21	AAV29134	pNF-kappab-cl2EGFP	813	16	1.2	4488	22	AAV09266	pT7-VPS vector DN
C 741	16	1.2	4172	20	AAV19901	Plasmid pOG277 DNA	814	16	1.2	4496	21	AAV53849	Vector pTGF82. S

815	16	1.2	4503	19	AAV13170	Complete DNA sequ
816	16	1.2	4515	14	AA046823	plISc-SF vector
817	16	1.2	4522	22	AAE26076	Feline IL-12p40 en
818	16	1.2	4522	22	AAE26076	Plasmid pXol-1112
819	16	1.2	4525	20	AAE69746	Nucleotide sequenc
820	16	1.2	4539	16	AA087347	Plasmid pINVI. Sy
821	16	1.2	4539	19	AAV37292	plINVI plasmid used
822	16	1.2	4540	18	AAE69891	Human cyclin D1-hu
823	16	1.2	4540	19	AAV54119	Human cyclin D1/cy
824	16	1.2	4554	21	AAE55541	atr reading frame
825	16	1.2	4557	18	AAE94591	Vector plasmid pw6
826	16	1.2	4570	22	AAE61152	Single chain Fv an
827	16	1.2	4574	19	AAV32664	Plasmid pCL25 stre
828	16	1.2	4583	10	AAAN90183	DNA sequence of pl
829	16	1.2	4583	12	AAO15174	Plasmid pAMVBT4.
830	16	1.2	4583	18	AAE59737	Vector pAMVBT5 enc
831	16	1.2	4583	19	AAE64256	Plasmid pPR7/8 DNA
832	16	1.2	4583	21	AAZ99198	Sequence of plasmid
833	16	1.2	4586	19	AAV3441	DNA sequence of th
834	16	1.2	4603	11	AAO04010	Plasmid pSUN387.
835	16	1.2	4608	9	AAAN81407	Sequence of new re
836	16	1.2	4608	10	AAAN90288	Plasmid pPR70-1.
837	16	1.2	4610	21	AAV5005	Nucleotide sequenc
838	16	1.2	4613	22	AAE59062	Plasmid pUCM5-6h73
839	16	1.2	4614	17	AAV5803	Vector pROCOS4/7.
840	16	1.2	4617	17	AAV72653	Rat tripeptidylpep
841	16	1.2	4618	21	AAZ39628	DNA sequence of pl
842	16	1.2	4620	22	AAZ39638	Vector pGN8 DNA se
843	16	1.2	4621	18	AAE59889	Human cyclin D1-hu
844	16	1.2	4621	19	AAV54117	Human cyclin D1/cy
845	16	1.2	4626	18	AAE79498	Plasmid pBS13. Sy
846	16	1.2	4626	22	AAE03988	pHSP-OROXDS/BH pla
847	16	1.2	4641	14	AAO40294	Sequence of plasmid
848	16	1.2	4643	13	AAO25112	DHFR-APP fusion pr
849	16	1.2	4643	22	AAZ06385	Vector pGX52 DNA s
850	16	1.2	4644	21	AAAS3850	Vector pTGC95. S
851	16	1.2	4645	21	AAAN92023	Recombinant plasmid
852	16	1.2	4648	12	AAO12756	Alpha-glucosidase/
853	16	1.2	4648	12	AAO12756	Alpha-glucosidase-
854	16	1.2	4651	10	AAE25044	Sequence of recomb
855	16	1.2	4655	13	AAO25108	DHFR-APP fusion pr
856	16	1.2	4657	19	AAV45804	Vector pROCOS5/3.
857	16	1.2	4660	10	AAAN90266	Plasmid pPRGIF2.
858	16	1.2	4668	11	AAO4519	Recombinant plasmid
859	16	1.2	4673	13	AAO25107	DHFR-APP fusion pr
860	16	1.2	4673	10	AAO25100	DHFR-APP fusion pr
861	16	1.2	4688	10	AAAN92029	Sequence of recomb
862	16	1.2	4689	21	AAZ49986	Plasmid pCG400 for
863	16	1.2	4690	13	AAO29910	RBI, PDI coexpress
864	16	1.2	4691	16	AAO32546	pCm3 expression
865	16	1.2	4694	22	AAE55225	Nucleotide sequenc
866	16	1.2	4701	22	AAAB9875	Plasmid pRT-L2. S
867	16	1.2	4701	22	AAAB9875	Vector pG218 DNA s
868	16	1.2	4713	19	AAV12067	Murine Iad alpha c
869	16	1.2	4715	10	AAAN0284	Plasmid pGFR2-15 c
870	16	1.2	4715	11	AAAN05083	Plasmid pSG1-12 of
871	16	1.2	4716	11	AAO05376	Portion of plasmid
872	16	1.2	4723	18	AAAT78802	Gamma heavy chain
873	16	1.2	4723	18	AAV93243	Plasmid pCG7-96 nu
874	16	1.2	4723	20	AAZ21997	Nucleotide sequenc
875	16	1.2	4723	22	AAE55224	Nucleotide sequenc
876	16	1.2	4724	19	AAV12068	Murine Iad beta ch
877	16	1.2	4724	21	AAE75924	Human ORFX ORF1479
878	16	1.2	4727	22	AAAD09265	PT7-TACS vector DN
879	16	1.2	4732	14	AAO03700	Recombinant plasmid
880	16	1.2	4732	14	AAO51730	Plasmid pRK5.1 for
881	16	1.2	4732	14	AAO51730	Nucleotide sequenc
882	16	1.2	4756	22	AAE06388	Vector pGX22 DNA s
883	16	1.2	4760	11	AAO25099	DHFR-APP fusion pr
884	16	1.2	4760	11	AAO05204	Portion of plasmid
885	16	1.2	4762	9	AAAN81675	Plasmid pBSF0EX1
886	16	1.2	4768	22	AAO50243	Plasmid vector pDG
887	16	1.2	4774	9	AAAN81424	Plasmid pBLAKI sta
888	16	1.2	4775	21	AAE56319	Cks-P29-Cks fusion
889	16	1.2	4776	20	AAE77617	Expression constru
890	16	1.2	4776	20	AAE77617	Expression constru
891	16	1.2	4789	9	AAAN82213	Plasmid pGIF1 enco
892	16	1.2	4789	9	AAAN82213	Plasmid pGIF1 enco
893	16	1.2	4816	22	AAO07799	PTCHS1B vector co
894	16	1.2	4816	22	AAO07799	Plasmid pPRAdelta
895	16	1.2	4819	15	AAO44360	Plasmid pPRAdelta
896	16	1.2	4819	18	AAE89768	Plasmid pPRAdelta
897	16	1.2	4819	18	AAE89768	Plasmid pPRAdelta
898	16	1.2	4819	18	AAE89768	Plasmid pPRAdelta
899	16	1.2	4819	18	AAE89768	Plasmid pPRAdelta
900	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
901	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
902	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
903	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
904	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
905	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
906	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
907	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
908	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
909	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
910	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
911	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
912	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
913	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
914	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
915	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
916	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
917	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
918	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
919	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
920	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
921	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
922	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
923	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
924	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
925	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
926	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
927	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
928	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
929	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
930	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
931	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
932	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
933	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
934	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
935	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
936	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
937	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
938	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
939	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
940	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
941	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
942	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
943	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
944	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
945	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
946	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
947	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
948	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
949	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
950	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
951	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
952	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
953	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
954	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
955	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
956	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
957	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
958	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
959	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
960	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
961	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
962	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
963	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
964	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
965	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
966	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
967	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
968	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
969	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
970	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
971	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
972	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
973	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
974	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
975	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
976	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
977	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
978	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
979	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
980	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
981	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
982	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
983	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
984	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
985	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
986	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
987	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
988	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
989	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
990	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
991	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
992	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
993	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
994	16	1.2	4824	22	AAE89768	Plasmid pPRAdelta
995	16					

Accession	Position	Score	Alignment
C 961	16	1.2	5015 18 AAT47449 Human somatogenic
C 962	16	1.2	5015 20 AAX87977 Human soluble grow
C 963	16	1.2	5015 20 AAY62766 Plasmid pJ1446 con
C 964	16	1.2	5015 21 AA288448 Plasmid pJ1446 enc
C 965	16	1.2	5021 21 AAA07775 DNA sequence of pl
C 966	16	1.2	5027 20 AAX78965 Human tissue vecto
C 967	16	1.2	5030 18 AAT75702 pGREEN1ANTEN-1 ve
C 968	16	1.2	5030 20 AAX08453 pPH1406 vector co
C 969	16	1.2	5033 15 AAO74695 Plasmid glucanyla
C 970	16	1.2	5042 16 AAO84694 Nucleotide sequenc
C 971	16	1.2	5051 22 AAC86954 Plasmid PCIB709
C 972	16	1.2	5056 10 AAN90795 Plasmid PSE-2 enco
C 973	16	1.2	5060 18 AAT78171 Plasmid PSE-2 enco
C 974	16	1.2	5062 18 AAT78170 Expression vector
C 975	16	1.2	5064 21 AAY18740 Complete DNA seque
C 976	16	1.2	5067 19 AAY18740 Human tissue facto
C 977	16	1.2	5069 20 AAX78871 Plasmid PKK232-8
C 978	16	1.2	5096 16 AAO83824 Plasmid pTat70PE
C 979	16	1.2	5098 15 AAO44355 Plasmid pTat70PE
C 980	16	1.2	5098 18 AAT89760 Plasmid pTat70PE
C 981	16	1.2	5098 18 AAT88225 Plasmid pTat70PE
C 982	16	1.2	5098 18 AAT84568 Plasmid pTat70PE
C 983	16	1.2	5098 19 AAV56584 DNA encoding B-HDA
C 984	16	1.2	5099 13 AAO24687 Nucleotide sequenc
C 985	16	1.2	5101 21 AAA10799 pTPCH12B vector co
C 986	16	1.2	5104 22 AAD07798 Rat brain homology
C 987	16	1.2	5109 17 AAT72654 Complete DNA seque
C 988	16	1.2	5109 19 AAY13846 Plasmid PCM-B64
C 989	16	1.2	5110 16 AAO83833 Plasmid PCM-P1208
C 990	16	1.2	5110 16 AAO83834 Plasmid PCM-P1031
C 991	16	1.2	5110 16 AAO83835 Plasmid PCM-T221
C 992	16	1.2	5110 16 AAO83836 Plasmid PCM-X#-S
C 993	16	1.2	5110 16 AAO83825 Plasmid PCM-T297
C 994	16	1.2	5110 16 AAO83827 Plasmid PCM-T297
C 995	16	1.2	5110 16 AAO83828 Plasmid PCM-T270
C 996	16	1.2	5110 16 AAO83828 Plasmid PCM-T270
C 997	16	1.2	5110 16 AAO83830 Plasmid PCM-P1087
C 998	16	1.2	5110 16 AAO83830 Plasmid PCM-P1087
C 999	16	1.2	5110 16 AAO83831 Plasmid PCM-P1198
C 1000	16	1.2	5110 16 AAO83832 Plasmid PCM-T286

ALIGNMENTS

RESULT 1

AAS00045 standard; DNA: 1362 BP.

AC AAS00045:

DT 14-MAY-2001 (first entry)

XX Human DNA encoding C2GnT3.

XX Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;

XX UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;

XX malignant melanoma; breast cancer; cervical cancer; hypodactylity;

XX hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;

XX Leukemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;

XX acquired immunodeficiency syndrome; sepsis; wound healing; infection; ds.

XX Homo sapiens.

XX OS

XX Key

XX CDS

XX primer_bind

XX primer_bind

XX primer_bind

/label= "Binding site for PCR primer TSHC121"

W0200114535-A2.
01-MAR-2001.
24-AUG-2000; 2000WO-DK00469.
24-AUG-1999; 99US-0150488.
(SCHW)/ SCHWIENTEK T.
(CLAU)/ CLAUSEN H.
Schwientek T, Clausen H;
WPI: 2001-226615/23.
P-PSDB; AAU00037.
New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
thymus-related disorders, cancers, tumours, immunosuppression

Claim 5; Fig 1; 97p; English.

The sequence encodes Human UDP-N-acetyl-glucosamine:Galactose-beta1,
3-N-acetylglucosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
(UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
preparation of compositions for treating a conditions mediated by C2GnT3,
particularly a thymus-related disorder. C2GnT3, nucleic acids
encoding it and antibodies against it may also be used for in vitro
purposes related to scientific research. DNA synthesis and manufacture of
vectors, in the prognostic and diagnostic evaluation of conditions
associated with altered expression or activity of C2GnT3 or conditions
requiring modulation of C2GnT3, as well as in monitoring conditions by
detecting and localising the DNA and protein. Disorders such as tumours
(e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
breast or cervix), hypodactylity, hyperactivity, atrophy, enlargement of
thymus, autoimmunity, arthritis, leukemia, lymphomas, immunosuppression,
acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
sepsis, wound healing, acute and chronic infection, cell-mediated or
humour immunity, or Th1/Th2 imbalance, may be treated using these protein
or nucleic acid. The antibodies may be used to screen potential
therapeutic compounds to determine their effects on a conditions such as
thymus-related disorder or cancer, to determine the level of C2GnT3
expression in cells genetically engineered to produce C2GnT3, or to
detect and quantify polypeptides in a sample to determine their role in a
particular cellular events or pathological states and to diagnose and
treat such pathological states.

Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match

100.0%; Score 1362; DB 22; Length 1362;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Accession	Position	Score	Alignment
QY	1	1362	1 atgaagatattcaaatgttattttaacatacctcagcagaagtttcatctgtt 60
DB	1	1362	1 atgaagatattcaaatgttattttaacatacctcagcagaagtttcatctgtt 60
QY	61	1362	61 ttaaccctatgctcgtctctcttctgttaagcttcaaatgtagaagcctttccgcaa 120
DB	61	1362	61 ttaaccctatgctcgtctctcttctgttaagcttcaaatgtagaagcctttccgcaa 120
QY	121	1362	121 aaagcatttacttggttgagbactccctaagtaacctgcctttgttaagaacagatac 180
DB	121	1362	121 aaagcatttacttggttgagbactccctaagtaacctgcctttgttaagaacagatac 180
QY	181	1362	181 actcatgttaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaag 240
DB	181	1362	181 actcatgttaagatgaagatgaagatgaagatgaagatgaagatgaagatgaagatgaag 240

```

QY 241 ccttggaaattgaaagagctggaataagaagaaggagacatcattgacttgaggat 300
    |||||
Db 241 ccttggaaattgaaagagctggaataagaagaaggagacatcattgacttgaggat 300
QY 301 gatgatgtgtggaatgacagcagtgatgtgacattatcagaactcctaaggatagct 360
    |||||
Db 301 gatgatgtgtggaatgacagcagtgatgtgacattatcagaactcctaaggatagct 360
QY 361 caaaagctgtctcaaggagagaaagcttcccaatgagcattcttctgtgtccac 420
    |||||
Db 361 caaaagctgtctcaaggagagaaagcttcccaatgagcattcttctgtgtccac 420
QY 421 aagatgcaattatgttgaagagctatccatgctataacacaccagacatatattac 480
    |||||
Db 421 aagatgcaattatgttgaagagctatccatgctataacacaccagacatatattac 480
QY 481 tgcatacatatgatgtgaagcagctgtaacctgaagtgccatgagcaatattagct 540
    |||||
Db 481 tgcatacatatgatgtgaagcagctgtaacctgaagtgccatgagcaatattagct 540
QY 541 aagtgcttcccaatatttcatgtgtccaaattagaggctgtggaataatgccacat 600
    |||||
Db 541 aagtgcttcccaatatttcatgtgtccaaattagaggctgtggaataatgccacat 600
QY 601 tccagactcagagctgatttcaattgtgtgagaccccttgcgaagcttgcgaatccag 660
    |||||
Db 601 tccagactcagagctgatttcaattgtgtgagaccccttgcgaagcttgcgaatccag 660
QY 661 aaatgttatcgaactgtgtggaagagatcttccctgaagcgaatttggatggct 720
    |||||
Db 661 aaatgttatcgaactgtgtggaagagatcttccctgaagcgaatttggatggct 720
QY 721 tcagagtgtaaaaaaactcaatggaagaataatgttggaagcgggtgaaccccaacag 780
    |||||
Db 721 tcagagtgtaaaaaaactcaatggaagaataatgttggaagcgggtgaaccccaacag 780
QY 781 aaatggaagatcattacatcagactgagacgggtgacctgaatgaatgtggaag 840
    |||||
Db 781 aaatggaagatcattacatcagactgagacgggtgacctgaatgaatgtggaag 840
QY 841 ctaccaaataagacaacatctccaaaggaagcaccocccataacattcagatattgt 900
    |||||
Db 841 ctaccaaataagacaacatctccaaaggaagcaccocccataacattcagatattgt 900
QY 901 ggcagctctatttgttgaagtcagacattgttaaatatttccaacatccatc 960
    |||||
Db 901 ggcagctctatttgttgaagtcagacattgttaaatatttccaacatccatc 960
QY 961 gttcaagacttttgcctgtgtctaaagacacatactctctgatgagcatttggct 1020
    |||||
Db 961 gttcaagacttttgcctgtgtctaaagacacatactctctgatgagcatttggct 1020
QY 1021 accctgttggtgtccaggaataacctggggaagatttccagatcagcccgagtgct 1080
    |||||
Db 1021 accctgttggtgtccaggaataacctggggaagatttccagatcagcccgagtgct 1080
QY 1081 gatctgagagtaagactcgcctgtgcaagtggaaattactactgaagcttttctatcc 1140
    |||||
Db 1081 gatctgagagtaagactcgcctgtgcaagtggaaattactactgaagcttttctatcc 1140
QY 1141 agtgtgactgatactcactcctcgaaagcgtgtgattatgagactcagagattaaagt 1200
    |||||
Db 1141 agtgtgactgatactcactcctcgaaagcgtgtgattatgagactcagagattaaagt 1200
QY 1201 ctatcaaaagatgacactgtgtgcaataaattgatactcaagtgagccctatctg 1260
    |||||
Db 1201 ctatcaaaagatgacactgtgtgcaataaattgatactcaagtgagccctatctg 1260
QY 1261 attaatgtcttgagcaaaagccttgaagaacagcagagagactgagatcttgcctca 1320
    |||||
Db 1261 attaatgtcttgagcaaaagccttgaagaacagcagagagactgagatcttgcctca 1320
QY 1321 gaaagttattatgataagaatctcactacacatcatga 1362

```

```

Db 1321 gaaagttattatgataagaatctcactacacatcatga 1362

RESULT 2
ID AAH98678 standard; cDNA; 781 BP.
AC AAH98678;
DT 12-OCT-2001 (first entry)
XX
XX Human EST-derived coding sequence SEQ ID NO: 535.
DE
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KM biodiversity; gene therapy; nutrition; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
PD
XX
XX 25-JAN-2001; 2001WO-US02687.
PF
XX
XX 25-JAN-2000; 2000US-0491404.
PR
XX 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-061451.
PR 15-SEP-2000; 2000US-0663870.
XX
XX (HXXE-) HXXEQU INC.
PA
XX
XX Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi Y;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX DR MPI: 2001-476164/51.
XX
XX P-PSDB; AMM24019.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
PR
XX
XX Claim 1; Page 549; 1275pp; English.
PS
XX
XX The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
SQ
XX
XX Sequence 781 BP; 218 A; 175 C; 174 G; 214 T; 0 other;

Query Match
Best Local Similarity 100.0%; Pred. No. 1,3e-157;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgaagattccaatgttattttaaacatcacacagcagaaagtttcatcctgtt 60
    |||||
Db 230 atgaagattccaatgttattttaaacatcacacagcagaaagtttcatcctgtt 289
QY 61 ttaacctgtgctcctccttctgttaagcttcaaatgtgaagcagctcttccgcaa 120
    |||||
Db 290 ttaacctgtgctcctccttctgttaagcttcaaatgtgaagcagctcttccgcaa 349
QY 121 aaagacattactgtgtgagactccctaagtaactcgccttgtgtaagaacagatac 180
    |||||
Db 350 aaagacattactgtgtgagactccctaagtaactcgccttgtgtaagaacagatac 409

```



```

QY 181 actcatgttaagatgaagtcgaagtaagaattacgttcgggtatcattgaacaaggag 240
    |||||||
Db 410 actcatgttaagatgaagtcgaagtaagaattacgttcgggtatcattgaacaaggag 469
QY 241 cctttggaattggaagatgctctggaataagaagaaggacatcattgacttggagat 300
    |||||||
Db 470 cctttggaattggaagatgctctggaataagaagaaggacatcattgacttggagat 529
QY 301 gatgatgttggtgcaatgaccagtgatgtgacattatcagaactctaa 349
    |||||||
Db 530 gatgatgttggtgcaatgaccagtgatgtgacattatcagaactctaa 578

```

RESULT 3

```

AACI9209
ID AACI9209 standard; CDNA; 186 BP.
XX
AC AACI9209;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 23284.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PE 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 990S-0122487.
XX
PA (GENSET) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 23284; 71bp + CD-ROM; English.
XX

```

The present sequence is one of a large number of 5' ESTs derived from cDNAs encoding secreted proteins. No ORF has yet been conclusively identified within the present sequence. The 5' ESTs were prepared from total human RNAs or poly(A)⁺ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 186 BP; 54 A; 37 C; 38 G; 57 T; 0 other;

Query Match

Best Local Similarity 13.4%; Score 183; DB 21; Length 186;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 60 tttaacctatgctgctctctcttggtaaagcttcaaatgtagagacactcttcogca 119
    |||||||
Db 1 tttaacctatgctgctctctcttggtaaagcttcaaatgtagagacactcttcogca 60

```

```

QY 120 aaaaagcattactctggttgagtactccctaagtactcctcgcttttgaagaacaagata 179
    |||||||
Db 61 aaaaagcattactctggttgagtactccctaagtactcctcgcttttgaagaacaagata 120
QY 180 cactcatgttaagatgaagtcgaagtaagaattacgttcgggtatcattgaacaagata 239
    |||||||
Db 121 cactcatgttaagatgaagtcgaagtaagaattacgttcgggtatcattgaacaagata 180
QY 240 gcc 242
    |||
Db 181 gcc 183

```

RESULT 4

```

AAS00048
ID AAS00048 standard; DNA; 31 BP.
XX
AC AAS00048;
XX
DT 14-MAY-2001 (first entry)
XX
DE Human C2GNT3 PCR primer TSHC100.
XX
KW Human; C2GNT3; Thymus-related disorder; cancer; tumour; adenoma;
KM UDP-GlcNAc; Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KM malignant melanoma; breast cancer; cervical cancer; hypoadalylty;
KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KM leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
KM TSHC100; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO200114535-A2.
XX
PR 01-MAR-2001.
XX
PE 24-AUG-2000; 2000WO-DK00469.
XX
PR 24-AUG-1999; 990S-0150488.
XX
PA (SCHW/) SCHWIENTEK T,
PA (CLAU/) CLAUSEN H.
XX
PI Schiewentek T, Clausen H;
XX
DR WPI; 2001-226615/23.
XX
PT New C2GNT3 polypeptides and nucleic acids encoding the polypeptides
PT useful for treating conditions mediated by a C2GNT3 polypeptide, e.g.,
PT thymus-related disorders, cancers, tumours, immunosuppression -
XX
PS Example 2; Page 56; 97pp; English.
XX

```

The sequence is PCR primer TSHC100 used to isolate DNA encoding amino acids 39-453 of Human UDP-N-acetyl-glucosamine:Galactose-beta1,3-N-acetylgalactosamine-alpha-R-beta1-6-N-acetylglucosaminyltransferase (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or C2GNT3). C2GNT3 and nucleic acids encoding it are useful in the preparation of compositions for treating a conditions mediated by C2GNT3, particularly a thymus-related disorder. C2GNT3, nucleic acids encoding it and antibodies against it may also be used for in vitro purposes related to scientific research. C2GNT3, nucleic acids encoding it, in the prognostic and diagnostic evaluation of conditions associated with altered expression or activity of C2GNT3 or conditions regulating modulation of C2GNT3, as well as in monitoring conditions by detecting and localising the DNA and protein. Disorders such as tumours (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the breast or cervix), hypoadalylty, hyperactivity, atrophy, enlargement of thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression, acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome, sepsis, wound healing, acute and chronic infection, cell-mediated or

CC humour immunity, or TH1/TH2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer to determine the level of C2GnT3
 CC expression in cells genetically engineered to produce C2GnT3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.
 CC
 SQ Sequence 31 BP; 10 A; 6 C; 7 G; 8 T; 0 other;
 Query Match 1.8%; Score 25; DB 22; Length 31;
 Best Local Similarity 100.0%; Pred. No. 0.04;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 114 tcgcgcaaaaagacattacttggtt 138
 Db 7 tcgcgcaaaaagacattacttggtt 31
 RESULT 5
 AAS00046/c
 ID AAS00046 standard; DNA; 21 BP.
 XX
 AC AAS00046;
 XX
 DT 14-MAY-2001 (first entry)
 DE Human C2GnT3 PCR primer TSHC96.
 XX
 KM Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
 KM UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
 KM malignant melanoma; breast cancer; cervical cancer; hypoadactivity;
 KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
 KM leukaemia; lymphoma; immunosuppression; AIDS; Miskott-Aldrich syndrome;
 KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
 KM TSHC96; PCR primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200114535-A2.
 XX
 PD 01-MAR-2001.
 XX
 PF 24-AUG-2000; 2000WO-DK00469.
 XX
 PR 24-AUG-1999; 99US-0150488.
 XX
 PA (SCHW/) SCHWIENTER T.
 PA (CLAU/) CLAUSEN H.
 XX
 PI Schwientek T, Clausen H;
 XX
 DR WPI; 2001-226615/23.
 XX
 PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 XX
 PS Example 1; Page 55; 97pp; English.
 XX
 CC The sequence represents PCR primer TSHC96 used to isolate genomic
 CC DNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 CC 3-N-acetylglactosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
 CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 CC preparation of compositions for treating a conditions mediated by C2GnT3,
 CC particularly a thymus-related disorder. C2GnT3, nucleic acids
 CC encoding it and antibodies against it may also be used for in vitro
 CC purposes related to scientific research, DNA synthesis and manufacture of
 CC vectors, in the prognostic and diagnostic evaluation of conditions
 CC associated with altered expression or activity of C2GnT3 or conditions

CC requiring modulation of C2GnT3, as well as in monitoring conditions by
 CC detecting and localising the DNA and protein. Disorders such as tumours
 CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 CC breast or cervix), hypoadactivity, hyperactivity, atrophy, enlargement of
 CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 CC acquired immunodeficiency syndrome (AIDS), Miskott-Aldrich syndrome,
 CC sepsis, wound healing, acute and chronic infection, cell-mediated or
 CC humour immunity, or TH1/TH2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer, to determine the level of C2GnT3
 CC expression in cells genetically engineered to produce C2GnT3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.
 CC
 SQ Sequence 21 BP; 5 A; 7 C; 3 G; 6 T; 0 other;
 Query Match 1.5%; Score 21; DB 22; Length 21;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 750 tatgttgagacggtggaacc 770
 Db 21 Tatgttgagacggtggaacc 1
 RESULT 6
 AAS00047
 ID AAS00047 standard; DNA; 32 BP.
 XX
 AC AAS00047;
 XX
 DT 14-MAY-2001 (first entry)
 DE Human C2GnT3 sequencing primer TSHC99.
 XX
 KM Human: C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
 KM UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
 KM malignant melanoma; breast cancer; cervical cancer; hypoadactivity;
 KM hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
 KM leukaemia; lymphoma; immunosuppression; AIDS; Miskott-Aldrich syndrome;
 KM acquired immunodeficiency syndrome; sepsis; wound healing; infection;
 KM TSHC99; primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200114535-A2.
 XX
 PD 01-MAR-2001.
 XX
 PF 24-AUG-2000; 2000WO-DK00469.
 XX
 PR 24-AUG-1999; 99US-0150488.
 XX
 PA (SCHW/) SCHWIENTER T.
 PA (CLAU/) CLAUSEN H.
 XX
 PI Schwientek T, Clausen H;
 XX
 DR WPI; 2001-226615/23.
 XX
 PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 XX
 PS Example 1; Page 55; 97pp; English.
 XX
 CC The sequence represents sequencing primer TSHC99 used to sequence a
 CC cDNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 CC 3-N-acetylglactosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
 CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or

CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 CC preparation of compositions for treating a conditions mediated by C2GnT3,
 CC particularly a thymus-related disorder. C2GnT3, nucleic acids
 CC encoding it and antibodies against it may also be used for in vitro
 CC purposes related to scientific research, DNA synthesis and manufacture of
 CC vectors, in the prognostic and diagnostic evaluation of conditions
 CC associated with altered expression or activity of C2GnT3 or conditions
 CC requiring modulation of C2GnT3, as well as in monitoring conditions by
 CC detecting and localising the DNA and protein. Disorders such as tumours
 CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 CC breast or cervix), hyporeactivity, hyperactivity, atrophy, enlargement of
 CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 CC sepsis, wound healing, acute and chronic infection, cell-mediated or
 CC humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer, to determine the level of C2GnT3
 CC expression in cells genetically engineered to produce C2GnT3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.

XX SQ Sequence 32 BP; 13 A; 4 C; 7 G; 8 T; 0 other;

Query Match 1.5%; Score 20; DB 22; Length 32;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaagatattcaattgtta 20
 ||||||||||||||||
 Db 13 atgaagatattcaattgtta 32

RESULT 7

AA062335/c ID AA062335 standard; cDNA; 1086 BP.

XX AC AA062335;

XX DT 27-SEP-1994 (first entry)

XX DE Influenza type B HA gene IN89.

XX KM Virus; vaccine; strain; diagnosis; prevention; ss.

XX OS Influenza virus type B.

XX FH Key Location/Qualifiers

XX FT CDS 46..1086

XX PN WO9407533-A.

XX PD 14-APR-1994.

XX PF 27-SEP-1993; 93WO-0509164.

XX PR 30-SEP-1992; 92US-0954840.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Rota PA;

XX DR WPI; 1994-135232/16.

XX P-PSDB; AAR51846.

XX New nucleotide sequences from influenza B virus strains - used to

XX PT develop prods. for preventing or diagnosing infection or for

XX PT identifying Influenza B virus strains

XX PS Claim 19; Page 41; 98pp; English.

XX The sequence is one of a number of influenza type B strain HA genes
 CC which produce a protein which can be used in a vaccine to induce a
 CC protective immune response to influenza type B in a mammal. The
 CC gene can be used for the prevention or diagnosis of influenza type B
 CC infection or for the identification of a strain of influenza type B.
 CC See also AA062334-48.

XX SQ Sequence 1086 BP; 387 A; 249 C; 225 G; 225 T; 0 other;

Query Match 1.5%; Score 20; DB 15; Length 1086;

Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 763 gTgaacccccaacagtaa 782
 ||||||||||||||||
 Db 627 GTGAACCCCAACAGTAA 608

RESULT 8

AAA31543 ID AAA31543 standard; DNA; 364 BP.

XX AC AAA31543;

XX DT 05-JUL-2000 (first entry)

XX DE Plant microsatellite marker #504.

XX Plant microsatellite sequence; core repeat sequence; detection; probe;
 KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KM variety identification; genetic variability evaluation; primer; ss.

XX OS Eucalyptus grandis.

XX PN WO967421-A1.

XX PD 29-DEC-1999.

XX PF 25-JUN-1999; 99WO-N200092.

XX PR 25-JUN-1998; 98US-0105307.

XX PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.

XX PI (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Havukala IJ, Bloksberg LN, Glenn M;

XX DR WPI; 2000-116958/10.

XX PT New plant microsatellite markers and associated flanking species for

XX PT the detection of polymorphic genetic markers -

XX PS Claim 1; Page 225; 392pp; English.

XX Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms. In genome mapping,
 CC identification and in positional cloning of genes, in variety
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

XX SQ Sequence 364 BP; 106 A; 71 C; 97 G; 89 T; 1 other;

Query Match 1.4%; Score 19; DB 21; Length 364;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 87 aaagcttcaatgcgaga 105
|||||
Db 163 aaagcttcaatgcgaga 181

RESULT 9
AAC41094
ID AAC41094 standard; DNA; 1127 bp.
XX
AC AAC41094;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 30623.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP103405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130444.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0133048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 03-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0133263.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

```

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149932.
PR 23-AUG-1999; 99US-0149933.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158829.
PR 08-OCT-1999; 99US-0158832.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 1.4%; Score 19; DB 21; Length 1127;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 295 gagatgatgatgtgtgtg 313
Db 94 gagatgatgatgtgtgtg 112

```

```

RESULT 10
AAS00049
ID AAS00049 standard; DNA; 20 BP.
XX
XX AAS00049;
AC
AC 14-MAY-2001 (first entry)
XX
XX Human C2GnT3 PCR primer TSHC101.
DE
DE Human; C2GnT3; Thymus-related disorder; cancer; tumour; adenoma;
XX
XX UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
XX malignant melanoma; breast cancer; cervical cancer; hypoadrenia;
XX hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
XX leukaemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
XX acquired immunodeficiency syndrome; sepsis; wound healing; infection;
XX TSHC101; PCR primer; ss.
XX
XX Homo sapiens.
XX
XX W0200114535-A2.
XX
XX 01-MAR-2001.
XX
XX 24-AUG-2000; 2000WO-DK00469.
XX
XX 24-AUG-1999; 99US-0150488.
XX
XX (SCHW/) SCHWIENTEK T.
XX (CLAU/) CLAUSEN H.
XX
XX Schwientek T, Clausen H;
XX
XX WPI; 2001-226615/23.
XX
XX New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
XX useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
XX thymus-related disorders, cancers, tumours, immunosuppression.
XX
XX Example 1; Page 55; 97pp; English.
XX
XX The sequence represents PCR primer TSHC101 used to isolate genomic
XX DNA encoding Human UDP-N-acetyl-glucosamine:Galactose-beta1,
XX 3-N-acetylglucosamine-alpha-R beta1-6 N-acetylglucosaminyltransferase
XX (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
XX C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
XX preparation of compositions for treating a conditions mediated by C2GnT3,
XX particularly a thymus-related disorder. C2GnT3, nucleic acids
XX encoding it and antibodies against it may also be used for in vitro
XX purposes related to scientific research, DNA synthesis and manufacture of
XX vectors, in the prognostic and diagnostic evaluation of conditions
XX associated with altered expression or activity of C2GnT3 or conditions
XX requiring modulation of C2GnT3, as well as in monitoring conditions by
XX detecting and localising the DNA and protein. Disorders such as tumours
XX (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
XX breast or cervix), hypoadrenia, atrophy, enlargement of
XX thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
XX acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
XX sepsis, wound healing, acute and chronic infection, cell-mediated or
XX humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
XX or nucleic acid. The antibodies may be used to screen potential
XX therapeutic compounds to determine their effects on a conditions such as
XX thymus-related disorder or cancer, to determine the level of C2GnT3
XX expression in cells genetically engineered to produce C2GnT3, or to
XX detect and quantify polypeptides in a sample to determine their role in a
XX particular cellular events or pathological states and to diagnose and
XX treat such pathological states.
XX
XX Sequence 20 BP; 5 A; 5 C; 4 G; 6 T; 0 other;

```

Query Match 1.3%; Score 18; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 495 tcgttaagcactgtatc 512
|||||
Db 1 tcgttaagcactgtatc 18

RESULT 11
AAH81918
ID AAH81918 standard; DNA; 255 BP.
AC AAH81918;
XX
XX 21-SEP-2001 (first entry)
DE Rat differential transcription-associated CDNA SEQ ID 427.
XX
XX Differential transcription; human; rat; tumour cell; cytostatic;
KM Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
XX Rattus sp.
OS
XX WO200157058-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 31-JAN-2001; 2001WO-EP01003.
PF
XX 31-JAN-2000; 2000DE-1004102.
PR
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
PA
XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;
PI Grips M, Hellriegel M, Schmitz A, Sers C;
PI WPI; 2001-483415/52.
DR
XX
XX Nucleic acids differentially expressed between tumor and normal cells,
PT useful for diagnosis or therapy of tumors and for screening active
PT agents
PS Claim 6; Page 495; 579pp; German.
XX
XX This invention describes a nucleic acid (I) with differential expression
CC between tumour and normal cells and which has cytostatic activity. (I)
CC work as modulators of Ras activity by inducing expression of tumour
CC suppressor genes. (I), and polypeptides encoded by them, are useful as
CC targets for diagnosis or therapy and in screening to determine the
CC effects of an active compound (potential pharmaceutical) on a cell line,
CC particularly for diagnosis and treatment of tumors, especially by
CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
CC methods) or by modulating the amount and/or location of (I)-encoded
CC polypeptides (by administration of the polypeptide or its activator,
CC antibody (optionally as a conjugate) or inhibitor). The method allows
CC identification of many Class II tumour suppressor genes (i.e. genes that
CC are not primary targets for tumour-initiating mutations).
CC AAH81492-AAH83376 represent the human and rat derived nucleic acid
CC fragments described in the method of the invention.
XX
XX Sequence 255 BP; 89 A; 57 C; 44 G; 65 T; 0 other;
SQ

Query Match 1.3%; Score 18; DB 22; Length 255;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1276 gaaagcttgaagacag 1293
|||||
Db 237 gaaagcttgaagacag 254

RESULT 12

AAZ13168/c
ID AAZ13168 standard; CDNA; 300 BP.
XX
XX AAZ13168;
AC
XX 12-OCT-1999 (first entry)
DE
XX
XX Human gene expression product CDNA sequence SEQ ID NO:637.
DE
XX
XX Human; gene: gene expression product; diagnosis; therapy; probe;
KM detection; mapping; tissue typing; profiling; forensic; cancer;
KM genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
OS
XX WO938972-A2.
PN
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99WO-US01619.
PF
XX
XX 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Crikvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Imis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat T;
PI Lamson G, Leszkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Grain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
DR
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
PT
PS Claim 1; Page 807-808; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX Sequence 300 BP; 71 A; 87 C; 62 G; 68 T; 12 other;
SQ

Query Match 1.3%; Score 18; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1003 gatgacactttggct 1020

Db 269 GATGAGACTTTGGCT 252

RESULT 13
AAFI1803/c
ID AAFI1803 standard; cDNA; 528 BP.

XX AAFI1803:

XX 13-MAR-2001 (first entry)

XX Aspergillus niger EST SEQ ID NO:4326.

XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX culture condition; environmental stress; spore morphogenesis;
XX metabolic pathway engineering; catabolic pathway engineering; ss.

XX Aspergillus niger.

XX WO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -

XX Claim 87; Page 1862; 3161pp; English.

XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stresses, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random cDNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAFI1803 to AAFI1803 represents ESTs from
XX Fusarium venenatum; AAFI1803 to AAFI1803 represents ESTs from
XX niger; AAFI1803 to AAFI1803 represents ESTs from Aspergillus
XX AFI14879 to AAFI1537 represents ESTs from Trichoderma reesei, which are
XX all specifically claimed in the present invention.

XX Sequence 528 BP; 129 A; 129 C; 135 G; 135 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 528;
Best Local Similarity 100.0%; Pred. No. 82;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 465 CCAGCACAATATTACTG 482
Db 326 CCAGCACAATATTACTG 309

RESULT 14

XX AA212009
XX ID AA212009 standard; DNA; 592 BP.

XX AA212009;

XX 08-OCT-1999 (first entry)

XX Neisseria meningitidis partial ORF17 sequence.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.

XX Neisseria meningitidis.

XX WO924578-A2.

XX 20-MAY-1999.

XX 09-OCT-1998; 98WO-IB01665.

XX 01-SEP-1998; 98GB-0019016.

XX 06-NOV-1997; 97GB-0022516.

XX 14-NOV-1997; 97GB-0024190.

XX 18-NOV-1997; 97GB-0024386.

XX 27-NOV-1997; 97GB-0025158.

XX 10-DEC-1997; 97GB-0026147.

XX 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

XX Grandi G, Masiagnan V, Pizsa M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

XX P-PSDB; AAY38541.

XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection

XX Claim 9; Page 101; 524pp; English.
XX Nucleotide sequences AA211972-212358 represent open reading frames
XX (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
XX antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
XX fragments, their nucleic acids and antibodies are used for diagnosis,
XX prevention (as vaccines) or treatment of Neisseria infections,
XX such as meningitis, septicemia and gonorrhea. Both organisms
XX are closely related. Fragments of the nucleic acids are useful
XX as hybridisation probes and antisense reagents.

XX Sequence 592 BP; 124 A; 190 C; 130 G; 145 T; 3 other;

Query Match 1.3%; Score 18; DB 20; Length 592;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 ttcacccgttttcaac 66
Db 128 ttcacccgttttcaac 145

RESULT 15

XX AA81331
XX ID AA81331 standard; DNA; 592 BP.

XX

```

AC  AAA81331;
XX
XX  04-DEC-2000 (first entry)
DT
XX
XX  N. meningitidis MenB polynucleotide sequence ORF number 10.
DE
XX  Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX  antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX  Meningococcus B; MenB; ds.
XX
XX  Neisseria meningitidis.
OS
XX  WO200022430-A2.
XX
XX  20-APR-2000.
XX
XX  08-OCT-1999; 99WO-US23573.
XX
XX  09-OCT-1998; 98US-0103794.
XX  30-APR-1999; 99US-0132068.
XX
XX  (CHIR ) CHIRON CORP.
XX
XX  -Prazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
XX  Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
XX  Rappuoli R, Pizza M.
XX  WPI: 2000-318079/27.
XX
XX  Isolated nucleotide sequences of Neisseria meningitidis which can be
XX  used in the diagnosis and treatment of N. meningitidis infection and
XX  other Neisserial infections, for example, N.gonorrhoea -
XX
XX  Disclosure: Page 202; 1760pp; English.
XX
XX  The present invention describes methods of obtaining immunogenic
XX  proteins from Neisseria genomic sequences. AAA81453 to AAA82414
XX  represent specifically claimed Neisseria meningitidis genomic DNA
XX  sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX  Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX  AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX  isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX  AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
XX  sequences, which are all used in the exemplification of the present
XX  invention. The nucleic acid sequences, protein sequences, and antibodies
XX  against them, can be used in the manufacture of a composition. The
XX  composition can be used as a medicament (or in the manufacture of a
XX  medicament) for treating, preventing or diagnosing infection due to
XX  Neisserial bacteria. For example, some of the identified proteins could
XX  be components of vaccines against Meningococcus B; against all serotypes;
XX  and/or against all pathogenic Neisseriae. Identification of sequences
XX  from the bacterium will also facilitate production of biological probes,
XX  particularly organism-specific probes. Attempts to make efficacious
XX  Meningococcus B vaccines have failed mainly due to antigen tolerance.
XX  Multivalent vaccines have also been tried but none have successfully
XX  overcome antigenic variability. The provision of further, complete
XX  sequences may provide an opportunity to identify secreted or surface
XX  exposed proteins that may be presumed targets for the immune system and
XX  which are not antigenically variable or at least more conserved than
XX  other more variable regions.
XX
XX  Sequence 592 BP; 124 A; 190 C; 130 G; 145 T; 3 other;
SO

```

```

Query Match      1.3%; Score 18; DB 21; Length 592;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  49 ttcatcctgttttaacc 66
    |||||||
DB  128 ttcatcctgttttaacc 145

```

```

RESULT 16
AAZ12010
ID  AAZ12010 standard; DNA; 807 BP.
XX
XX  AAZ12010;
AC
XX  08-OCT-1999 (first entry)
DT
XX
XX  Complete ORF17 sequence of Neisseria meningitidis.
DE
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX  treatment; Neisseria infection; meningitis; septicemia; gonorrhoea; ss.
XX
XX  Neisseria meningitidis.
OS
XX  WO9924578-A2.
XX
XX  20-MAY-1999.
XX
XX  09-OCT-1998; 98WO-IB01665.
XX
XX  01-SEP-1998; 98GB-0019016.
XX  06-NOV-1997; 97GB-0023516.
XX  14-NOV-1997; 97GB-0024150.
XX  18-NOV-1997; 97GB-0024386.
XX  27-NOV-1997; 97GB-0025158.
XX  10-DEC-1997; 97GB-0026147.
XX  14-JAN-1998; 98GB-0000759.
XX
XX  (CHIR-) CHIRON SPA.
XX
XX  Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V,
XX  P-PSDB; AAY38542.
XX
XX  WPI: 1999-327407/27.
XX
XX  Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX  diagnosis, treatment and prevention of infection
XX
XX  Claim 9; Page 102; 524pp; English.
XX
XX  Nucleotide sequences AAZ11972-Z12358 represent open reading frames
XX  (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
XX  antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
XX  fragments, their nucleic acids and antibodies are used for diagnosis,
XX  prevention (as vaccines) or treatment of Neisseria infections,
XX  such as meningitis, septicemia and gonorrhoea. Both organisms
XX  are closely related. Fragments of the nucleic acids are useful
XX  as hybridisation probes and antisense reagents.
XX
XX  Sequence 807 BP; 153 A; 252 C; 192 G; 209 T; 1 other;
SO

```

```

Query Match      1.3%; Score 18; DB 20; Length 807;
Best Local Similarity 100.0%; Pred. No. 81.
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  49 ttcatcctgttttaacc 66
    |||||||
DB  343 ttcatcctgttttaacc 360

```

```

RESULT 17
AAZ12011
ID  AAZ12011 standard; DNA; 807 BP.
XX
XX  AAZ12011;
AC
XX  08-OCT-1999 (first entry)
DT
XX
XX  Complete ORF17 sequence of Neisseria meningitidis strain A.
XX  Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX

```



```

KM treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX
XX Neisseria meningitidis.
OS
XX WO924578-A2.
XX
XX 20-MAY-1999.
XX
XX
XX 09-OCT-1998; 98WO-IB01655.
XX
XX 01-SEP-1998; 98GB-0019016.
XX
XX 06-NOV-1997; 97GB-0023516.
XX
XX 14-NOV-1997; 97GB-0024190.
XX
XX 18-NOV-1997; 97GB-0024386.
XX
XX 27-NOV-1997; 97GB-0025158.
XX
XX 10-DEC-1997; 97GB-0026147.
XX
XX 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
PI WPI; 1999-327407/27.
XX
XX P-PSDB; AAY38543.
XX
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
XX
XX Claim 9; Page 103; 524pp; English.
XX
XX Nucleotide sequences AA211972-212358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX
XX
XX Sequence 807 BP; 152 A; 252 C; 193 G; 210 T; 0 other;
SO

Query Match 1.3%; Score 18; DB 20; Length 807;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ttcacccgtttttaacc 66
   ||||||||||||||||
Db 343 ttcacccgtttttaacc 360

RESULT 18
AA212012
ID AA212012 standard; DNA; 807 BP.
XX
XX AA212012;
AC
XX
XX 08-OCT-1999 (first entry)
XX
XX Complete ORF17 sequence of Neisseria gonorrhoeae.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM treatment; Neisseria infection; meningitis; septicemia; gonorrhea; ss.
XX
XX Neisseria meningitidis.
OS
XX
XX WO924578-A2.
XX
XX 20-MAY-1999.
XX
XX 09-OCT-1998; 98WO-IB01665.
XX
XX 01-SEP-1998; 98GB-0019016.
XX
XX
XX

```

```

PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
XX (CHIR-) CHIRON SPA.
XX
XX Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
PI WPI; 1999-327407/27.
XX
XX P-PSDB; AAY38545.
XX
XX
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
XX
XX Claim 9; Page 104; 524pp; English.
XX
XX Nucleotide sequences AA211972-212358 represent open reading frames
CC (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC fragments, their nucleic acids and antibodies are used for diagnosis,
CC prevention (as vaccines) or treatment of Neisseria infections,
CC such as meningitis, septicemia and gonorrhea. Both organisms
CC are closely related. Fragments of the nucleic acids are useful
CC as hybridisation probes and antisense reagents.
XX
XX
XX Sequence 807 BP; 147 A; 246 C; 199 G; 215 T; 0 other;
SO

Query Match 1.3%; Score 18; DB 20; Length 807;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ttcacccgtttttaacc 66
   ||||||||||||||||
Db 343 ttcacccgtttttaacc 360

RESULT 19
AAT93294
ID AAT93294 standard; DNA; 1160 BP.
XX
XX AAT93294;
AC
XX
XX 27-APR-1998 (first entry)
XX
XX Tomato mottle virus AC1 open reading frame.
DE
XX Geminivirus; TOMOV; AC1 gene; transdominant mutation;
KM transgenic plant; disease resistance; ss; cyclic; circular.
XX
XX Tomato mottle virus isolate Florida.
OS
XX
XX Key Location/Qualifiers
FH 44..1129
FT /*tag= a
XX
XX WO9739110-A1.
XX
XX 23-OCT-1997.
XX
XX 15-APR-1997; 97WO-US06300.
XX
XX 16-APR-1996; 96US-0015517.
XX
XX (SEMI-) SEMINIS VEGETABLE SEEDS INC.
XX (WISC-) WISCONSIN ALUMNI RES FOUND.
XX
XX Ahlquist PG, Hanson SF, Liu HT, Maxwell DP, Stout JT;
PI WPI; 1997-526447/48.
XX
XX
XX

```

DR P-PSDB; AAW34332.
 XX Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Example 3.3; Page 57-58; 132pp; English.
 XX
 CC This genomic DNA sequence includes the open reading frame of the
 CC wild-type AC1 gene of tomato mottle virus (TOMOV), a geminivirus
 CC that has a bipartite genome. The AC1 gene must be expressed for
 CC efficient replication of the two genomic components, DNA-A and
 CC DNA-B. It encodes a protein (see AAW34336) having a DNA binding
 CC site specific to the DNA-A common region, a DNA nicking activity,
 CC and an NTP binding activity. The invention involves production of
 CC transgenic plants containing DNA comprising AC1 or CI wild-type or
 CC mutant sequences that negatively interfere in trans with
 CC geminiviral replication during infection. Such transgenic plants
 CC are resistant to viral infection. The AC1/CI genes are especially
 CC from TOMOV, tomato yellow leaf curl virus or bean golden mosaic
 CC geminivirus (see AAT93282-93) and encode polypeptides (see AAW34324-35)
 CC that have mutations in the highly conserved DNA-nicking domain
 CC and/or the NTP-binding domain.
 CC
 SO Sequence 1160 BP; 360 A; 276 C; 257 G; 267 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1160;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagattccagatc 1064
 ||||||||||||||||
 DB 380 tggggagattccagatc 397
 RESULT 20
 AAT93285
 ID AAT93285 standard; DNA; 1166 BP.
 AC AAT93285;
 XX 27-APR-1998 (first entry)
 DE Tomato mottle virus AC1 mutant gene.
 XX
 KM Geminivirus; TOMOV-AC1d1m23; AC1 gene; transdominant mutation;
 KM transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 44..439
 FT /*tag- a
 FT
 XX WO9739110-A1.
 PN 23-OCT-1997.
 PD 23-OCT-1997.
 PF 15-APR-1997; 97WO-US06300.
 PR 16-APR-1996; 96US-0015517.
 PR 16-APR-1996; 96US-0015517.
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX WPI; 1997-526447/48.
 DR P-PSDB; AAW34337.
 XX

PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Claim 11; Page 72-73; 132pp; English.
 XX
 CC This DNA sequence comprises a mutated AC1 gene of tomato mottle
 CC virus (TOMOV). It carries an inserted 4-base Sau3A site that
 CC shifts the AC1 gene translation reading frame resulting in the
 CC expression of a truncated AC1 protein (see AAW34327). The AC1 gene
 CC (see also AAT93294) must be expressed for efficient replication of
 CC the two genomic components, DNA-A and DNA-B, of the bipartite TOMOV
 CC genome. The invention involves production of transgenic plants
 CC containing DNA comprising geminivirus AC1 or CI wild-type or mutant
 CC sequences that negatively interfere in trans with geminiviral
 CC replication during infection. Such transgenic plants are resistant
 CC to viral infection. The AC1/CI genes are especially from TOMOV,
 CC tomato yellow leaf curl virus or bean golden mosaic geminivirus
 CC (see AAT93282-93) and encode polypeptides (see AAW34324-35) that have
 CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
 CC domains.
 CC
 SO Sequence 1166 BP; 361 A; 277 C; 260 G; 268 T; 0 other;
 Query Match 1.3%; Score 18; DB 18; Length 1166;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagattccagatc 1064
 ||||||||||||||||
 DB 380 tggggagattccagatc 397
 RESULT 21
 AAT93282
 ID AAT93282 standard; DNA; 1169 BP.
 AC AAT93282;
 XX 27-APR-1998 (first entry)
 DE Tomato mottle virus AC1 mutant TOMOV-AC1d1m gene.
 XX
 KM Geminivirus; TOMOV-AC1d1m; AC1 gene; transdominant mutation;
 KM transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 44..1129
 FT /*tag- a
 FT
 XX WO9739110-A1.
 PN 23-OCT-1997.
 PD 23-OCT-1997.
 PF 15-APR-1997; 97WO-US06300.
 PR 16-APR-1996; 96US-0015517.
 PR 16-APR-1996; 96US-0015517.
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
 XX WPI; 1997-526447/48.
 DR P-PSDB; AAW34324.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection

PN	MO9739110-A1.
XX	
PD	23-OCT-1997.
XX	
PF	15-APR-1997; 97WO-US06300.
XX	
PR	16-APR-1996; 96US-0015517.
XX	
PA	(SEMI-) SEMINIS VEGETABLE SEEDS. INC.
PA	(WISC) WISCONSIN ALUMNI RES FOUND.
XX	
PI	Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
DR	WPI: 1997-526447/48.
DR	P-PSDB; AAM34325.
XX	
PT	Transgenic plants expressing geminivirus AC1 and C1 wild-type and
PT	mutant genes - have increased resistance to geminivirus infection
PT	e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
PT	golden mosaic geminivirus

FN	MO9739110-A1.
XX	
PD	23-OCT-1997.
XX	
PE	15-APR-1997; 97WO-US06300.
XX	
PR	16-APR-1996; 96US-0015517.
XX	
PA	(SEMI-) SEMINIS VEGETABLE SEEDS. INC.
PA	(MISC) WISCONSIN ALUMNI RES FOUND.
XX	
PI	Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT;
XX	
DR	WPI; 1997-526447/48.
DR	P-PSDB; AAM34326.
XX	
PT	Transgenic plants expressing geminivirus AC1 and CI wild-type and
PT	mutant genes - have increased resistance to geminivirus infection
PT	e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
PT	golden mosaic geminivirus
XX	
PS	Claim 11; Page 67-69; 132pp; English.

XX This DNA sequence comprises a transdominant lethal mutant.
CC designated TOMOV-AC1d1m23, of the AC1 gene of tomato mottle virus
CC virus (TOMOV). It encodes an AC1 protein (see AAW4326) that carries
CC 2 mutations in an NTP-binding domain. The AC1 gene (see also
CC AAT93294) must be expressed for efficient replication of the two
CC genomic components, DNA-A and DNA-B, of the bipartite TOMOV genome.
CC The invention involves production of transgenic plants containing
CC DNA comprising geminivirus AC1 or CI wild-type or mutant sequences
CC that negatively interfere in trans with geminiviral replication
CC during infection. Such transgenic plants are resistant to viral
CC infection. The AC1/CI genes are especially from TOMOV, tomato
CC yellow leaf curl virus or bean golden mosaic geminivirus (see
CC AAT93282-93) and encode polypeptides (see AAW34324-35) that have
CC mutations in the highly conserved DNA-nicking and/or the NTP-binding
CC domains.
XX
SQ Sequence 1169 BP; 364 A; 278 C; 257 G; 270 T; 0 other;

Query Match 1.3%; Score 18; DB 18; Length 1169;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1047 tggggagattccagatc 1064
|||||
Db 380 tggggagattccagatc 397

RESULT 24
AAC42240/C
ID AAC42240 standard; DNA; 1463 BP.
XX
AC AAC42240;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 34807.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136352.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145341.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151063.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```

```

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161930.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 18; DB 21; Length 1463;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 474 tattactgcattcatta 491
Db 750 TATTACTGCATTCATTA 733

RESULT 25
ID AAA81548 standard; DNA; 2456 BP.
AC AAA81548;
XX 04-DEC-2000 (first entry)
DT N. meningitidis partial DNA sequence gnm_95 SEQ ID NO:95.
DE N. meningitidis partial DNA sequence gnm_95 SEQ ID NO:95.
XX KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX KM Meningococcus B; Menb; ds.
OS Neisseria meningitidis.
XX PN WO200022430-A2.
XX PD 20-APR-2000.
XX PF 08-OCT-1999; 99WO-US23573.
XX PR 09-OCT-1999; 98US-0103794.
XX PR 30-APR-1999; 99US-0132068.
XX RA (CHIR) CHIRON CORP.
XX PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappuoli R, Pizza M;
XX DR WPI: 2000-318079/27.
XX PT Isolated nucleotide sequences of Neisseria meningitidis which can be
XX PT used in the diagnosis and treatment of N. meningitidis infection and
XX PT other Neisserial infections, for example, N.gonorrhoea -
XX PS Claim 7; Page 1487-1488; 1760pp; English.
XX CC The present invention describes methods of obtaining immunogenic
XX CC proteins from Neisseria genomic sequences. AAA8153 to AAA8244
XX CC represent specifically claimed Neisseria meningitidis genomic DNA
XX CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
XX CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
XX CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
XX CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
XX CC AAA81452 represent Neisseria meningitidis Menb polynucleotide ORF
XX CC sequences, which are all used in the exemplification of the present
XX CC invention. The nucleic acid sequences, protein sequences, and antibodies
XX CC against them, can be used in the manufacture of a composition. The
XX CC composition can be used as a medicament (or in the manufacture of a

```

medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B, against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 CC
 SQ Sequence 2456 BP; 553 A; 753 C; 565 G; 585 T; 0 other;
 CC
 Query Match 1.3%; Score 18; DB 21; Length 2456;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 49 ttcatccgttttacc 66
 DB 1805 ttcatccgttttacc 1822
 CC
 RESULT 26
 AA193309/C
 ID AA193309 standard; DNA; 2602 BP.
 AC
 XX AA193309;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Tomato mottle virus full-length A-component clone.
 XX
 KW Geminivirus; TOMOV; AC1 gene; transdominant mutation;
 KW transgenic plant; disease resistance; ss; cyclic; circular.
 XX
 OS Tomato mottle virus isolate Florida.
 XX
 PN WO9739110-A1.
 XX
 PD 23-OCT-1997.
 XX
 PF 15-APR-1997; 97WO-US06300.
 XX
 PR 16-APR-1996; 960S-0015517.
 XX
 PA (SEMI-) SEMINIS VEGETABLE SEEDS INC.
 XX (MISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Ahlquist PG, Hanson SF, Luu HT, Maxwell DP, Stout JT.
 DR WPI: 1997-526447/48.
 XX
 PT Transgenic plants expressing geminivirus AC1 and CI wild-type and
 PT mutant genes - have increased resistance to geminivirus infection
 PT e.g. tomato mottle virus, tomato yellow leaf curl virus or bean
 PT golden mosaic geminivirus
 XX
 PS Example 3.1; Page 76-77; 132pp; English.
 XX
 CC This genomic DNA sequence comprises a full-length A-component clone
 CC of tomato mottle virus (TOMOV), a geminivirus that has a bipartite
 CC genome comprising DNA-A and DNA-B. It was isolated from TOMOV
 CC infected Nicotiana benthamiana and tomato plant DNA by restriction
 CC digestion. The DNA-B component (see AA193310) was also isolated.
 CC TOMOV DNA-A contains the AC1 gene (see AA193294) that must be
 CC expressed for efficient replication of DNA-A and DNA-B. The
 CC invention involves production of transgenic plants containing DNA
 CC comprising AC1 or CI wild-type or mutant sequences that negatively
 CC interfere in trans with geminiviral replication during infection.

Such transgenic plants are resistant to viral infection.
 CC
 SQ Sequence 2602 BP; 671 A; 561 C; 586 G; 784 T; 0 other;
 CC
 Query Match 1.3%; Score 18; DB 18; Length 2602;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1047 tggggagattccagc 1064
 DB 2273 TGGGAGATTTCAGATC 2256
 CC
 RESULT 27
 AA257860/C
 ID AA257860 standard; cDNA; 2843 BP.
 AC
 XX AA257860;
 XX
 DT 11-APR-2000 (first entry)
 XX
 DE Protein regulating gene expression PRGE-22 cDNA clone 2998209.
 XX
 KW Protein regulating gene expression; PRGE-22; human;
 KW cancer; inflammation; anticancer; antitumor; antiinflammatory;
 KW zinc finger; apoptosis inhibitor; antiapoptotic; gene therapy;
 KW diagnosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Location/Qualifiers
 FT CDS 377..1822
 FT /*tag= a
 XX
 PN WO9964596-A2.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US13281.
 XX
 PR 12-JUN-1998; 98US-0089029.
 PR 29-JUL-1998; 98US-0094575.
 PR 14-OCT-1998; 98US-0104624.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;
 PI Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DM;
 DR WPI: 2000-116543/10.
 DR P-PSDB: AA158629.
 XX
 PT New human polypeptides that regulate gene expression, for treatment,
 PT prevention and diagnosis of, e.g. cancer.
 XX
 PS Claim 9; Page 138; 150pp; English.
 XX
 CC The present sequence is that of Incyte clone 2998209 encoding new
 CC human protein regulating gene expression PRGE-22 (see AA158629). The
 CC cDNA was initially isolated from ovarian tumor tissue cDNA library
 CC OVAR107, and the full-length sequence assembled from overlapping
 CC clones from a number of libraries. PRGE-22 is expressed in
 CC reproductive, nervous and gastrointestinal tissues associated with
 CC cancer, inflammation and foetal diseases, disorders or conditions.
 CC It is characterised as a zinc finger protein and inhibitor of
 CC apoptosis. The invention provides PRGE polypeptides (see AA158608-38)
 CC and polynucleotides (see AA257839-69), expression vectors, host cells,
 CC antibodies, agonists and antagonists. It also provides methods for
 CC diagnosing, treating or preventing disorders associated with
 CC expression of PRGE. Polynucleotides are also used as sources of
 CC probes and primers for diagnosis and monitoring of disease, also
 CC for detecting related sequences and in gene mapping.

XX Sequence 2843 BP; 835 A; 537 C; 662 G; 809 T; 0 other;
 SQ Query Match 1.3%; Score 18; DB 21; Length 2843;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 ttctacccctgtttttaa 64
 ||||||||||||||||
 Db 476 TTTTCATCCGTGTTTAA 459

RESULT 28
 AAF44664
 ID AAF44664 standard; cDNA: 3864 BP.
 XX
 AC AAF44664;
 XX
 DT 27-MAR-2001 (first entry)

XX Novel protein kinase cDNA, SEQ ID NO: 44.
 XX

XX Human; mouse; protein kinase; antiarthritic; atherosclerotic; osteopathic;
 KM immunosuppressive; cardiant; renal; antiinflammatory; antiaschematic;
 KM dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KM immune disorder; cardiovascular disease; neurodegenerative disease;
 KM cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KM inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

XX Homo sapiens.
 OS
 PN WO200073469-A2.
 XX
 PD 07-DEC-2000.

XX 28-MAY-2000; 2000MO-US14842.
 XX
 PR 28-MAY-1999; 99US-0136503.
 XX

PA (SUGEN-) SUGEN INC.
 XX

PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
 XX

DR WPI; 2001-032161/04.
 DR
 P-PSDB: AAB65637.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
 PT treating immune-related diseases and disorders, cardiovascular disease,
 PT neurodegenerative diseases and/or cancers -
 XX
 XX Example 1; Fig 2; 310pp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids
 CC and the protein kinases they encode may be used in the treatment and
 CC diagnosis of diseases associated with inappropriate kinase expression
 CC such as immune-related diseases and disorders, cardiovascular disease,
 CC neurodegenerative diseases and/or cancers. The nucleic acids and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays. The kinase polypeptides may be used as antigens in the production
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies
 CC and kinase antagonists may also be used to down regulate kinase
 CC expression and activity. Diseases related to kinase expression and
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
 CC disorders, complications of organ transplantation, myocardial infarction,
 CC immune disorders, cardiomyopathies, strokes, renal failure,
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
 CC reproductive disorders.

SQ Sequence 3864 BP; 1048 A; 995 C; 995 G; 826 T; 0 other;

XX Query Match 1.3%; Score 18; DB 22; Length 3864;
 SQ Best Local Similarity 100.0%; Pred. No. 76;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 tcatgacttgaggatg 301
 ||||||||||||||||
 Db 3368 tcatgacttgaggatg 3385

RESULT 29
 AAC75062/c
 ID AAC75062 standard; cDNA: 4982 BP.
 XX
 AC AAC75062;
 XX
 DT 08-FEB-2001 (first entry)

XX Human ORF517 polynucleotide sequence SEQ ID NO:1233.
 XX

XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 KM vulnerable; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KM antiautismic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.

XX Homo sapiens.
 OS
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 PI
 DR WPI; 2000-602362/57.
 DR
 P-PSDB: AAB40853.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX Claim 5; Page 1091-1094; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; antiautismic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 4982 BP; 1475 A; 901 C; 1096 G; 1508 T; 2 other;

Query Match 1.3%; Score 18; DB 21; Length 4982;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 ttttcacgtgttttaa 64
 ||||||||||||||||
 Db 1189 ttttcacgtgttttaa 1172

RESULT 30

AX89220
 ID AX89220 standard; DNA; 5355 BP.

AC AX89220;

DT 15-SEP-1999 (first entry)

DE Seq ID No: 28 of WO9933961.

KM DRAK1: DAP kinase related apoptosis inducing kinase; human;
 KM apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
 KM viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
 KM Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
 KM diabetes; ss.

XX Homo sapiens.

XX WO9933961-A1.

PD 08-JUL-1999.

PF 25-DEC-1998; 98WO-JP05974.

PR 17-APR-1998; 98JP-0108150.

PR 26-DEC-1997; 97JP-0367640.

PR 26-DEC-1997; 97JP-0367641.

PR 17-APR-1998; 98JP-0108149.

XX (ASAH) ASAH KASEI KOGYO KK.

XX Akira S, Kawai T;

XX WPI, 1999-430239/36.

DR P-PSDB; AAY7163.

PT New kinase with apoptosis induction activity useful in the treatment
 of cancer, autoimmune diseases and viral infections

XX Disclosure; Page 157-166; 180pp; Japanese.

XX The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
 CC apoptosis inducing kinase) having apoptosis inducing activity. The
 CC kinases can be expressed recombinantly by transforming host cells with
 CC vectors comprising the nucleic acids encoding the kinases. The kinases
 CC are useful in the treatment, prevention, diagnosis and investigation of
 CC diseases with which apoptosis is associated, such as hormonally regulated
 CC cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
 CC diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
 CC Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
 CC rheumatoid arthritis; and diabetes.

XX Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;

Query Match 1.3%; Score 18; DB 20; Length 5355;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 284 tcatgacttgagagatg 301
 ||||||||||||||||
 Db 3491 tcatgacttgagagatg 3508

RESULT 31

AX89221/C
 ID AX89221 standard; DNA; 5355 BP.

AC AX89221;

DT 15-SEP-1999 (first entry)

DE Seq ID No: 30 of WO9933961.

KM DRAK1: DAP kinase related apoptosis inducing kinase; human;
 KM apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
 KM viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
 KM Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
 KM diabetes; ss.

XX Homo sapiens.

XX WO9933961-A1.

PD 08-JUL-1999.

PF 25-DEC-1998; 98WO-JP05974.

PR 17-APR-1998; 98JP-0108150.

PR 26-DEC-1997; 97JP-0367640.

PR 26-DEC-1997; 97JP-0367641.

PR 17-APR-1998; 98JP-0108149.

XX (ASAH) ASAH KASEI KOGYO KK.

XX Akira S, Kawai T;

XX WPI, 1999-430239/36.

PT New kinase with apoptosis induction activity useful in the treatment
 of cancer, autoimmune diseases and viral infections

XX Disclosure; Page 173-176; 180pp; Japanese.

XX The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
 CC apoptosis inducing kinase) having apoptosis inducing activity. The
 CC kinases can be expressed recombinantly by transforming host cells with
 CC vectors comprising the nucleic acids encoding the kinases. The kinases
 CC are useful in the treatment, prevention, diagnosis and investigation of
 CC diseases with which apoptosis is associated, such as hormonally regulated
 CC cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
 CC diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
 CC Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
 CC rheumatoid arthritis; and diabetes.

XX Sequence 5355 BP; 1260 A; 1302 C; 1278 G; 1515 T; 0 other;

Query Match 1.3%; Score 18; DB 20; Length 5355;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 284 tcatgacttgagagatg 301
 ||||||||||||||||

Db 1865 TCATTGACTTGAGGATG 1848

RESULT 32

AAZ39404
ID AAZ39404 standard; cDNA to mRNA; 5355 BP.

AC AAZ39404;

DT 23-FEB-2000 (first entry)

XX Human Trad protein encoding nucleotide sequence.

XX Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW skeletal muscle; human; ss.

OS Homo sapiens.

PN JPL1290083-A.

PD 26-OCT-1999.

PF 17-APR-1998; 98JP-0108151.

PR 17-APR-1998; 98JP-0108151.

PA (ASAH) ASAH KASEI KOGYO KK.

DR WPI; 2000-016982/02.

DR P-PSDB; AAY56781.

PT A new protein Trad or its salts - useful for the screening of diseases
involving the Rho subfamily on the function of skeletal muscles

PS Claim 4; Page 13-18; 25pp; Japanese.

CC This DNA encodes a human protein Trad. The protein is useful for the
screening and the evaluation of an inducer or an inhibitor of

CC phosphoenzymatic activity of Trad and for the diagnosis of diseases
participated by Rho subfamily on the function of skeletal muscles.

CC Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 5355;

Best Local Similarity 100.0%; Pred. No. 75;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcattgacttgaggatg 301
|||||

Db 3491 tcattgacttgaggatg 3508

RESULT 33

AAZ39405/C
ID AAZ39405 standard; cDNA to mRNA; 5355 BP.

AC AAZ39405;

DT 23-FEB-2000 (first entry)

XX Human Trad protein DNA complementary nucleotide sequence.

XX Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW skeletal muscle; human; complementary; ss.

OS Homo sapiens.

PN JPL1290083-A.

PD 26-OCT-1999.

PF 17-APR-1998; 98JP-0108151.

XX 17-APR-1998; 98JP-0108151.

XX (ASAH) ASAH KASEI KOGYO KK.

DR WPI; 2000-016982/02.

PT A new protein Trad or its salts - useful for the screening of diseases
involving the Rho subfamily on the function of skeletal muscles

PS Claim 9; Page 19-20; 25pp; Japanese.

CC The invention provides a human protein Trad. The protein is useful for
the screening and the evaluation of an inducer or an inhibitor of

CC phosphoenzymatic activity of Trad and for the diagnosis of diseases
participated by Rho subfamily on the function of skeletal muscles. The

CC present sequence represents the complementary sequence of the Trad DNA.

CC Sequence 5355 BP; 1260 A; 1302 C; 1278 G; 1515 T; 0 other;

Query Match 1.3%; Score 18; DB 21; Length 5355;

Best Local Similarity 100.0%; Pred. No. 75;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 tcattgacttgaggatg 301
|||||

Db 1865 TCATTGACTTGAGGATG 1848

RESULT 34

AAD02701
ID AAD02701 standard; DNA; 27150 BP.

AC AAD02701;

DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfoltransferase-6 (GST-6) genomic DNA #1.

CC Human; glycosyl sulfoltransferase-6; GST-6; immunosuppressive;
therapy; selectin binding inhibitor; gene therapy; inflammation;

CC systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

CC glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenarthritis;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

CC demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

CC asthma; hypersensitivity; rheumatic fever; tissue rejection; ds.

OS Homo sapiens.

PN WO200106015-A1.

PD 25-JAN-2001.

PF 19-JUL-2000; 2000WO-US19741.

PR 20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

PA (REGC) UNTV CALIFORNIA.

PI Rosen SD, Lee JK, Hemmerlich S;

DR WPI; 2001-138471/14.

PT New glycosyl sulfoltransferases (GST)-4alpha, GST-4beta and GST-6 for
diagnostic and therapeutic agent screening applications

PS Example 2; Page 116-123; 128pp; English.

CC The present sequence is human glycosyl sulfoltransferase-6 (GST-6)


```

FT primer_bind /tag= j
FT 4582..4600
FT /tag= k
FT misc_binding /tag= l
FT 4589..4613
FT /note= "binds probe"
FT replace(4601,G)
FT /tag= m
FT primer_bind complement(4602..4620)
FT /tag= n
FT exon 4627..4718
FT /tag= o
FT /label= "B"
FT primer_bind 4891..4908
FT /tag= p
FT primer_bind 10007..10025
FT /tag= q
FT exon 10115..10233
FT /tag= r
FT /label= "C"
FT primer_bind 10209..10227
FT /tag= s
FT misc_binding 10216..10240
FT /tag= t
FT /note= "binds probe"
FT replace(10228,T)
FT /tag=
FT primer_bind complement(10229..10247)
FT /tag= v
FT primer_bind 10267..10285
FT /tag= w
FT misc_binding 10274..10298
FT /tag= x
FT /note= "binds probe"
FT replace(10286,T)
FT /tag= y
FT primer_bind complement(10287..10305)
FT /tag= z
FT replace(10370,)
FT /tag= aa
FT primer_bind 10411..10430
FT /tag= ab
FT exon 26810..26897
FT /tag= ac
FT /label= "D"
FT exon 31357..31471
FT /tag= ad
FT /label= "E"
FT exon 34261..34404
FT /tag= ae
FT /label= "F"
FT exon 37377..37466
FT /tag= af
FT /label= "S"
FT exon 39704..40858
FT /tag= ag
FT /label= "T"
FT primer_bind 39556..39574
FT /tag= ah
FT primer_bind 39877..39896
FT /tag= ai
FT primer_bind 39925..39943
FT /tag= aj
FT misc_binding 39932..39956
FT /tag= ak
FT /note= "binds probe"
FT replace(39944,T)
FT /tag= al
FT primer_bind complement(39945..39963)
FT /tag= am
FT primer_bind 39953..39970
FT /tag= an
FT primer_bind 39954..39972

FT misc_binding /tag= ao
FT 39961..39985
FT /tag= ap
FT /note= "binds probe"
FT replace(39973,C)
FT /tag= aq
FT primer_bind complement(39974..39992)
FT /tag= ar
FT primer_bind 40242..40259
FT /tag= as
FT primer_bind 41137..41154
FT /tag= at
FT primer_bind 41366..41384
FT /tag= au
FT misc_binding 41373..71397
FT /tag= av
FT /note= "binds probe"
FT replace(41385,C)
FT /tag= aw
FT primer_bind 41385..41403
FT /tag= ax
FT complement(41386..41404)
FT /tag= ay
FT /tag= az
FT /note= "binds probe"
FT replace(41404,C)
FT /tag= ba
FT complement(41405..41423)
FT /tag= bb
FT primer_bind 41564..41581
FT /tag= bc
FT primer_bind 42122..42141
FT /tag= bd
FT primer_bind 42213..42231
FT /tag= be
FT misc_binding 42220..42244
FT /tag= bf
FT /note= "binds probe"
FT replace(42232,C)
FT /tag= bg
FT primer_bind complement(42233..42251)
FT /tag= bh
FT primer_bind 42526..42543
FT /tag= bi
FT exon 50436..50545
FT /tag= bj
FT /label= "G"
FT primer_bind 67289..67309
FT /tag= bk
FT primer_bind 67456..67474
FT /tag= bl
FT misc_binding 67463..67487
FT /tag= bm
FT /note= "binds probe"
FT replace(67475,G)
FT /tag= bn
FT primer_bind complement(67476..67494)
FT /tag= bo
FT primer_bind 67724..67741
FT /tag= bp
FT primer_bind 69182..69200
FT /tag= bq
FT primer_bind 69502..69520
FT /tag= br
FT misc_binding 69509..69533
FT /tag= bs
FT /note= "binds probe"
FT replace(69521,G)
FT /tag= bt
FT primer_bind complement(69522..69540)
FT /tag= bu
FT primer_bind 69609..69626

```

FT	primer_bind	/*tag=	by
FT		72698..72715	
FT	primer_bind	/*tag=	bw
FT		72819..72837	
FT		/*tag=	bx
FT	misc_binding	72826..72850	
FT		/*tag=	by
FT		/note=	"binds probe"
FT	allele	replace(72838,'T')	
FT		/*tag=	bz
FT	primer_bind	complement(72839..72857)	
FT		72861..72918	
FT	exon	/*tag=	ca
FT		72881..72918	
FT		/*tag=	cd
FT		label=	"H"
FT		73099..73117	
FT	primer_bind	/*tag=	cc
FT			

```

Query Match 1.3%; Score 18; DB 22; Length 240825;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Gaps 0;
QY 1207 aaagatcgacattggtt 1224
|||||
b 90632 aaagatcgacattggtt 90649
|||||

```

CC	RESULT 37
CC	AA161371
CC	ID AA161371 standard; DNA: 335913 BP.
CC	AC AA161371;
CC	DT 16-OCT-2001 (first entry)
CC	DE Soybean 240017 region G3, SEQ ID NO: 2.
CC	KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;
CC	SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
CC	KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
CC	OS Glycine max.
CC	PN W0200151627-A2.
CC	PD 19-JUL-2001.
CC	XX 05-JAN-2001; 2001WO-US00552.
CC	XX 07-JAN-2000; 2000US-0174880.
CC	PR (MONS) MONSANTO CO.
CC	PA Haughe BM, Wang ML, Parsons JD, Parnell LD;
CC	PI WPI; 2001-425872/45.
CC	XX P-PSDB; AAM42214.
CC	DR
CC	XX
CC	PT New purified nucleic acid for producing a soybean plant having soybean
CC	PT cyst nematode resistance and for use in plant breeding programs -
CC	XX
CC	PS Claim 2; Page 204-400; 1353pp; English.
CC	XX
CC	XX The invention relates to nucleic acid molecules from regions of the
CC	CC soybean genome which are associated with soybean cyst nematode (SCN)
CC	CC resistance. The nucleic acids are used to transform plants, and can
CC	CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC	CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC	CC of soybean plants and for introgressing SCN resistance or partial SCN
CC	CC resistance into soybean plants. They can also be used in plant breeding
CC	CC programmes. The invention also relates to proteins encoded by such
CC	CC nucleic acid molecules, as well as antibodies capable of recognising
CC	CC these proteins. The present sequence is a nucleic acid molecule

CC provided in the specification.
XX
SQ Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 other;

Query Match	1.3%	Score 18;	DB 22;	Length 335913;
Best Local Similarity	100.0%	Pred. No. 63;		
Matches	18;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0

QY	293	tgaggagcatgcatgttgc	310
Db	125725	tgaggagcatgcatgttgc	125742

RESULT	38
AAI61372	
ID	AAI61372 standard; DNA; 335913 BP

```

...      16-OCT-2001  (first entry)
DT
XX      Soybean 240017 region G3, SEQ ID NO: 3
DE

```

KW1 Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.

OS	Glycine max.
XX	
PN	W0200151627-A2.

PD	19-JUL-2001.
XX	
PF	05-JAN-2001; 2001WO-US00552.

PR 07-JAN-2000; 2000US-0174880.
XX
PA (MONS) MONSANTO CO.

PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
DR WPI; 2001-425872/45.

XX	New purified nucleic acid for producing a soybean plant having soybean
PT	cyst nematode resistance and for use in plant breeding programs -
PT	

PS Claim 2; page 400-595; 1353pp; English
XX
CC The invention relates to nucleic acid

CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an *rhg1* or an *Rhg4* SCN resistant allele.
CC The nucleic acids can be used for investigating *rhg1* or *Rhg4* haplotypes

resistance into soybean plants. They can also be used in plant breeding programmes. The invention also relates to proteins encoded by such nucleic acid molecules, as well as antibodies capable of recognising

CC provided in the specification.
XX
SQ sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 other;

Query Match	1.3%	Score	18	DB	22	Length	335913
Best Local Similarity	100.0%	Pred. No.	63				
Matches	18	Conservative	0	Mismatches	0	Indels	0
				Gaps			0

```
QY      293  tgagagatgatgtgtg  310
          |||||
Db 125725 tgagagatgatgtgtg 125742
```

```
RESULT 39
AAAF21610
ID AAAF21610 standard; DNA: 349980 BP.
XX
AC AAAF21610;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarcelli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
XX
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
```

```
Query Match 1.3%; Score 18; DB 21; Length 349980;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 49 ttatctctgttttaacc 66
|||||
```

DB 137432 ttcacctgttttaacc 137449

```
RESULT 40
AAAV21209
ID AAAV21209 standard; DNA: 1664976 BP.
XX
AC AAAV21209;
XX
DT 10-NOV-1998 (first entry)
XX
DE Methanococcus jannaschii circular chromosome.
XX
KM Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KM genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS Methanococcus jannaschii.
XX
PN WO9807830-A2.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14900.
XX
PR 22-AUG-1996; 96US-0024428.
XX
PA (GENO-) INST GENOMIC RES.
PA (UNIT) UNIV ILLINOIS FOUND.
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Bult CJ, Smith HO, Venter JC, White OR, Moese CR;
PI WPI; 1998-169145/15.
XX
DR Complete genome sequence of methano-genic archaeon, Methanococcus
DR jannaschii - useful in identification of M. jannaschii genome
XX
PT fragment
XX
PS Claim 13; Page 152-585; 614pp; English.
XX
CC The present sequence represents the complete 1.66-megabase pair genome
CC sequence of the Methanococcus jannaschii circular chromosome. The
CC present invention describes M. jannaschii open reading frames from the
CC genome sequence. The invention also describes a computer based system
CC for identifying fragments of the M. jannaschii genome that are
CC homologous to target nucleotide sequences, comprising: (a) data storage
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC bp sequence (see AA21209, AA21210 and AA21211), or a nucleotide
CC sequence at least 99.9% identical to it; (b) search means for comparing a
CC target sequence to the nucleotide sequence of the data storage means to
CC identify a homologous sequence, and (c) retrieval means for obtaining
CC the homologous sequence. The method, which is based on whole genome
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC of which consists of 3 physically distinct elements, a large circular
CC chromosome (the 1664976 bp sequence given in AA21209), a large circular
CC extra-chromosomal element (the 58407 bp sequence given in AA21210), and
CC a small circular extra-chromosomal element (the 16550 bp sequence given
CC in AA21211), can be used in the identification of M. jannaschii genome
CC fragment.
XX
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;
```

```
Query Match 1.3%; Score 18; DB 19; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 721 tcagagttgaataaac 738
|||||
DB 146412 tcagagttgaataaac 146429
```

RESULT 41


```

XX OS Homo sapiens.
XX PN WO200016916-A2.
XX PD 06-APR-2000.
XX PF 23-SEP-1999; 99WO-05222226.
XX PR 28-SEP-1998; 98US-0102161.
XX PR 28-SEP-1998; 98US-0102180.
XX PR 29-SEP-1998; 98US-0102380.
XX PR 08-OCT-1998; 98US-0103815.
XX PR 27-OCT-1998; 98US-0105877.
XX PA (CHIR ) CHIRON CORP.
XX PA (HYSE-) HYSEQ INC.
XX PI Williams LT, Escobedo J, Tunis MA, Garcia PD, Sudduth-Klinger J;
XX PI Rainhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX PI Lamson G, Drmanac R, Ckrvenjakov R, Dickson M, Drmanac S, Labat I;
XX PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX PS Claim 1; Page 344; 502pp; English.
XX PT Polynucleotide library comprising 1079 defined sequences, useful in
XX PT the form of an array to detect cancer or susceptibility to cancer .
XX PS Claim 1; Page 344; 502pp; English.
XX CC The present invention describes a library of polynucleotides comprising
XX CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
XX CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
XX CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
XX CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
XX CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
XX CC of detecting differentially expressed genes correlated with a cancerous
XX CC state of a mammalian cell comprising detecting a gene product encoded by
XX CC one of the 1079 sequences given in the specification. The polynucleotides
XX CC are used to monitor patients having (or susceptible) to cancer to detect
XX CC potentially malignant events at a molecular level before they are
XX CC detectable at a gross morphological level. The polynucleotides are also
XX CC useful for monitoring the efficacy of various therapies and preventive
XX CC interventions. Polynucleotide probes based on the disclosed sequences
XX CC are useful for chromosome mapping and detection of transcription levels.
XX CC The 1079 polynucleotide sequences were derived from a human colon cancer
XX CC cell line Km12cl-A cDNA library.
XX SQ Sequence 376 BP; 94 A; 90 C; 113 G; 79 T; 0 other;
SQ Query Match 1.2%; Score 17; DB 21; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 755 tggagacggtgaacc 771
DB 104 TGGAGACGGTGAACCC 88

```

```

OS OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMANA GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX PI WPI: 2001-235357/24.
XX DR P-PSDB; AAG76056.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX PS Claim 1; Page 4128-4132; 9803pp; English.
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patient's own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated P,
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAG77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 553 BP; 155 A; 108 C; 134 G; 150 T; 6 other;
SQ Query Match 1.2%; Score 17; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 46 gtttcatcctgtttt 62
DB 130 gtttcatcctgtttt 146

```

```

RESULT 44
AAH35461
ID AAH35461 standard; cDNA; 553 BP.
AC AAH35461;
XX 03-SEP-2001 (first entry)
DE Human colon cancer antigen encoding cDNA SEQ ID NO:2543.
XX Human colon cancer antigen encoding cDNA; diagnosis: detection;
KW colorectal carcinoma; chromosome 17; ss.
XX

```

```

RESULT 45
AAI43142
ID AAI43142 standard; DNA; 575 BP.
AC AAI43142;
XX 17-OCT-2001 (first entry)
DE Probe #11828 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX

```

```

XX 30-JAN-2001; 2001WO-US00663.
PF
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11828; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 575 BP; 179 A; 87 C; 79 G; 230 T; 0 other;

Query Match          1.2%; Score 17; DB 22; Length 575;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 553 aatatttcattgtcttc 569
   |||||||
Db 24 aatatttcattgtcttc 40

RESULT 46
AAH61796/c
ID AAH61796 standard; DNA; 724 BP.
XX
XX AAH61796;
AC
XX
XX 19-JUL-1999 (first entry)
DT
XX
XX B. burgdorferi antigenic protein coding sequence; t4-15.nt.
DE
XX
XX Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
KM
XX
XX Borrelia burgdorferi.
OS
XX
XX W09859071-A1.
PN
XX
XX 30-DEC-1998.
PD
XX
XX 18-JUN-1998; 98WO-US12718.
PF
XX
XX 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX (MED-) MEDIMUNE INC.
XX
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
XX WPI; 1999-189980/16.
DR
XX P-PDB; AAY20099.
XX

```

```

PI New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
XX Claim 1; Page 196-197; 275pp; English.
XX
XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 724 BP; 347 A; 125 C; 85 G; 167 T; 0 other;

Query Match          1.2%; Score 17; DB 20; Length 724;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 tgtgtggcgaagatttc 693
   |||||||
Db 294 TGTGTGCGCAAGATTTC 278

RESULT 47
AAH67794
ID AAH67794 standard; DNA; 732 BP.
XX
XX AAH67794;
AC
XX
XX 26-SEP-2001 (first entry)
DT
XX
XX C glutamicum coding sequence fragment SEQ ID NO: 2823.
DE
XX
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
KM
XX
XX organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0158162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOWA) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR
XX
XX P-PDB; AAG92575.
DR
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 8; SEQ ID NO: 2823; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing

```


CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX
 SQ Sequence 732 BP; 168 A; 201 C; 197 G; 166 T; 0 other;

Query Match 1.2%; Score 17; DB 22; Length 732;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1273 gcagaaagcttgaaga 1289
 |||||
 Db 145 gcagaaagcttgaaga 161

RESULT 48

AA36834/C
 ID AAX36834 standard; DNA; 816 BP.

XX
 AC AAX36834;

XX
 DT 14-JUL-1999 (first entry)

XX
 DE Human XLIS gene fragment including exon 4.

XX
 KW XLIS gene; human; detection; diagnosis; prenatal diagnosis; therapy;
 KW lissencephaly; LIS; agyria-pachygyria; subcortical laminar heterotopia;
 KW SCH; cortical dysgenesis; cryptogenic epilepsy; neurological disorder;
 KW neurodegenerative disease; Alzheimer's disease; X-linked disorder;
 KW genetic counselling; ss.

XX
 OS Homo sapiens.

XX
 PN EP918091-A1.

XX
 PD 26-MAY-1999.

XX
 PF 21-NOV-1997; 97EP-0402811.

XX
 PR 21-NOV-1997; 97EP-0402811.

XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX
 PI Chelly J, Des Portes V, Kahn A, Pinard J;

XX
 DR WPI; 1999-290318/25.

XX
 PT New gene and its gene product expressed in the brain, useful for
 PT diagnosing and treating disorders such as lissencephaly and
 PT subcortical laminar heterotopia

XX
 PS Claim 1; Page 23; 71pp; English.

XX This sequence is a fragment of the human XLIS gene of the invention.
 CC The XLIS fragments may be used to detect abnormalities in the expression
 CC of the XLIS gene transcripts or to compare their sequence with that of
 CC the XLIS transcripts from patients for in vitro especially prenatal
 CC diagnosis of lissencephaly (LIS) (or agyria-pachygyria), subcortical
 CC laminar heterotopia (SCH), cortical dysgenesis, cryptogenic epilepsies
 CC or neurodegenerative diseases such as Alzheimer's disease. These
 CC disorders mainly affect females as the XLIS gene is X-linked. The XLIS
 CC fragments may also be used to administer to patients to prevent or treat
 CC the above disorders and may be used as a tool in genetic counselling.
 CC Oligonucleotides which bind to the fragments may be used to amplify the
 CC XLIS gene from a sample for comparison to normal samples in the in vitro
 CC diagnosis regime. This may also be performed by amplifying XLIS cDNA from
 CC the mRNA in the sample. Antibodies to XLIS may be used to detect XLIS in
 CC a biological sample or can be administered to patients to prevent or
 CC treat the above disorders. They may also be used to purify XLIS from a

CC biological sample. XLIS may also be administered to patients to prevent
 CC or treat the above neurological disorders. In addition XLIS may be used
 CC as a marker of neuronal cells at an early stage of development. Its
 CC discovery increases understanding of both the neuronal movement which
 CC leads to development of the cortical region of the brain and of the
 CC pathogenesis of the group of neuronal disorders mentioned above.

XX
 SQ Sequence 816 BP; 204 A; 170 C; 178 G; 262 T; 2 other;

Query Match 1.2%; Score 17; DB 20; Length 816;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 agagaaagctccca 396
 |||||
 Db 353 AGGAGAAAGCTCCCA 337

RESULT 49

AA61795/C
 ID AAX61795 standard; DNA; 837 BP.

XX
 AC AAX61795;

XX
 DT 19-JUL-1999 (first entry)

XX
 DE B. burgdorferi antigenic protein coding sequence, f4-15.nt.

XX
 KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX
 OS Borrelia burgdorferi.

XX
 PN W09859071-A1.

XX
 PD 30-DEC-1998.

XX
 PF 18-JUN-1998; 98WO-US12718.

XX
 PR 03-SEP-1997; 97US-0057483.

XX
 PR 20-JUN-1997; 97US-0050359.

XX
 PR 22-JUL-1997; 97US-0053344.

XX
 PR 22-JUL-1997; 97US-0053377.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 PI (MEDI-) MEDIMUNE INC.

XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX
 DR WPI; 1999-18980/16.

XX
 DR P-PSDB; AAY20098.

XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases
 PT caused by Borrelia, particularly Lyme disease

XX
 PS Claim 1; Page 196; 275pp; English.

XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus.

XX
 SQ Sequence 837 BP; 396 A; 134 C; 98 G; 209 T; 0 other;

Query Match 1.2%; Score 17; DB 20; Length 837;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 tctgtggaagattt 693
 Db 344 TGTGTGGCAAGATTTT 328

Db 336 ctaataattgattct 352

RESULT 50
 AAC59007
 ID AAC59007 standard; cDNA: 917 BP.
 XX

Search completed: May 2, 2002, 13:20:48
 Job time: 6189 sec

AAC59007;

02-FEB-2001 (first entry)

Human secreted protein coding sequence SEQ ID NO: 28.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnarary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.

OS Homo sapiens.

PN WO200055175-A1.

PD 21-SEP-2000.

PF 09-MAR-2000; 2000WO-US06049.

PR 12-MAR-1999; 99US-0124144.

PR 11-JUN-1999; 99US-0138574.

PR 03-DEC-1999; 99US-0168667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI; 2000-638175/61.

DR P-PSDB; AAB27577.

PS Claim 1; Page 353; 428bp; English.

XX The invention relates to the isolation of genes AAB27560-B27609. The genes can be used to
 CC 50 human secreted proteins AAB27560-B27609. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (SEQID) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

SO Sequence 917 BP; 281 A; 170 C; 166 G; 300 T; 0 other;

Query Match 1.2%; Score 17; DB 21; Length 917;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 ctaataattgattct 1242

Fri May 3 10:57:49 2002

us-09-645-192-1.0110.rng

Page 36

XX WO200114535-A2.
 PN
 XX 01-MAR-2001.
 PD
 XX 24-AUG-2000; 2000WO-DK00469.
 PF
 XX 24-AUG-1999; 99US-0150486.
 PR
 XX (SCHW/) SCHWIENTEK T.
 PA (CLAU/) CLAUSEN H.
 XX
 PI Schwiientek T, Clausen H;
 XX
 DR WPI: 2001-226615/23.
 DR N-PSDB; AAS00045.
 XX
 PT New C2GnT3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C2GnT3 polypeptide, e.g.,
 XX thymus-related disorders, cancers, tumours, immunosuppression
 XX
 PS Claim 17; Fig 1; 97pp; English.

The sequence represents Human UDP-N-acetyl-glucosamine:Galactose-beta1,
 CC 3-N-acetylglactosamine-alpha-R beta1-6-N-acetylglucosaminyltransferase
 CC (UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 CC C2GnT3). C2GnT3 and nucleic acids encoding it are useful in the
 CC preparation of compositions for treating a conditions mediated by C2GnT3,
 CC particularly a thymus-related disorder. C2GnT3, nucleic acids
 CC encoding it and antibodies against it may also be used for in vitro
 CC purposes related to scientific research, DNA synthesis and manufacture of
 CC vectors, in the prognostic and diagnostic evaluation of conditions
 CC associated with altered expression or activity of C2GnT3 or conditions
 CC requiring modulation of C2GnT3, as well as in monitoring conditions by
 CC detecting and localising the DNA and protein. Disorders such as tumours
 CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 CC breast or cervix), hypocoactivity, hyperactivity, atrophy, enlargement of
 CC thymus, autoimmunity, arthritis, leukaemia, lymphomas, immunosuppression
 CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 CC sepsis, wound healing, acute and chronic infection, cell-mediated or
 CC humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer to determine the level of C2GnT3
 CC expression in cells genetically engineered to produce C2GnT3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.
 CC
 XX
 XX
 SQ Sequence 453 AA;

Query Match 100.0%; Score 2389; DB 22; Length 453;
 Best Local Similarity 100.0%; Pred. No. 2.1e-207;
 Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTFKCYFKHLLQOKYFLFLTMTLSLKLVNRRLLPOKDIYEVSLSPFRNRY 60
 DB 1 MKTFKCYFKHLLQOKYFLFLTMTLSLKLVNRRLLPOKDIYEVSLSPFRNRY 60
 QY 61 THYKDEVREVNCSGIYEPEPLIGKSLERRRDIIDEDDDVYVNTSCDIDYQTLRGYA 120
 DB 61 THYKDEVREVNCSGIYEPEPLIGKSLERRRDIIDEDDDVYVNTSCDIDYQTLRGYA 120
 QY 121 QXIVSKEEKSEPTAYSLVYHKDAIMVERLHAHYNOHNHYCHYDKKADPTKVMANNNA 180
 DB 121 QXIVSKEEKSEPTAYSLVYHKDAIMVERLHAHYNOHNHYCHYDKKADPTKVMANNNA 180
 QY 181 KCSNFIASKEAVEYAHISRLQADUNCLSDLLKSSIQMKYVYNLCGDFPLKSNFELY 240
 DB 181 KCSNFIASKEAVEYAHISRLQADUNCLSDLLKSSIQMKYVYNLCGDFPLKSNFELY 240
 QY 241 SELKTLNGAMNLETVKPNKSKLERFTYHHELRVPEYVKLPRTNISKAPPHNIQIV 300

DB 241 selktlngamnlvtxpnsklertfthelrrvpyvklpirtnskaephnqilv 300
 QY 301 GSAYFVLQAFVKYIFNNNSIVQDFFAMSKDTPSPDEHFHATLLRVPISPEISRSADVS 360
 DB 301 gsayfvlqsatvkyilfnnsivqdfawskdypscdehfwatlirvpipelsrsadvs 360
 QY 361 DLOSKTRLVKWNYYEGFPFPCSTSHLRSCIVGAALRWLIDGHHFAKFDKVPIL 420
 DB 361 dlosktrlvkwnyyegfipscstshlrsvclgaaelrwlidgghfankfidskvdpl 420
 QY 421 IKCLAEKLEPOQDMITLPSKELFMDRLFTTS 453
 DB 421 ikclaekeegqrgwiltlpskelfmdrnlftts 453

RESULT 2
 AAB18995
 ID AAB18995 standard; Protein: 438 AA.
 XX
 AC AAB18995;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE A core 2 beta-1,6-N-acetylglucosaminyltransferase.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis.
 XX
 OS Homo sapiens.
 XX
 PN CA2296936-A1.
 PD 03-AUG-2000.
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX
 PA (GLYC-) GLYCDESIGN INC.
 XX
 PI Korczak B, Lew A;
 XX
 DR WPI: 2000-594746/57.
 DR N-PSDB; AAA96569.
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 PS Claim 3; Page 50-51; 66pp; English.

The present sequence represents a human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular

DT 06-MAR-2001 (first entry)

XX Amino acid sequence of beta-1-6-N-acetylglucosaminyltransferase.

DE Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;

XX membrane protein; branched sialyl lex; L-selectin; immune reaction;

KW inflammation; tissue rejection; tumour metastasis.

XX

OS Homo sapiens.

XX

EH Key Location/Qualifiers

FT Modified-site 289 /note= "potential N-glycosylation site"

XX

XX US6136580-A.

XX

PD 24-OCT-2000.

XX

XX 19-JAN-1999; 99US-0233506.

XX

XX 19-JAN-1999; 99US-0233506.

XX

XX (BURN-) BURNHAM INST.

XX

XX Fukuda M, Yeh J;

XX

XX WPI; 2001-040238/05.

DR N-PSDB; AAC62134.

XX

XX

PT New C2GNT-M polypeptides having core 2, core 4 and I branching

PT beta-1-6-N-acetylglucosaminyltransferase activities for preparing

PT reagents useful for diagnosing, preventing or treating inflammation or

PT tumour metastasis.

XX

XX

PS Claim 1; Fig 4; 25pp; English.

XX

CC The present sequence represents a human

CC beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4

CC and I branching activities. It is designated C2GNT-M. C2GNT-M is a

CC membrane protein that is predominantly expressed in colon, small

CC intestine, trachea, stomach and thyroid, as well as in certain cancer

CC cell lines. C2GNT-M polypeptides may be used to prepare molecules having

CC highly branched sialyl lex and L-selectins, which may be subsequently

CC used to modulate immune reactions, e.g. inflammation and tissue

CC rejection, and to prevent or inhibit tumour metastasis.

CC

XX

SQ Sequence 438 AA;

Query Match 36.3%; Score 868; DB 22; Length 438;

Best Local Similarity 43.0%; Pred. No. 5.6e-70;

Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

XX

QY 23 LWLISLIKILNV-----RLFPQKDIYVEYLSLSPFYRN-RYTHVDEVREYVNCSG 75

DB 13 lwaigcmilatvalklsfrlkcdshlglesresgsqycrnlllyflkprksincsg 72

QY 76 IY--EQLP--EIGKSLERIRDDIIDEEDDVAMTSDCIVYTLGAYQKIKSEKESF 131

DB 73 vtvgdgaavgalnmlevkkrr-epftdchylsiltdechnfaekrfjgfpisveef 131

QY 132 PIASLVVHKDAIWEFLIAITNOHNICYIHYDRKAPDTFKVAMNNLAKCFSNIFIAK 191

DB 132 playsmvlhekienferllrayvaypqnlycvhvdexspetfkaavkaiiscfpnvflaak 191

QY 192 LEAVYAHISRILOADLCLSDLLKSSIQMKYVINIGODPILKSNFEIYSELKLNKGANV 251

DB 192 lrvvyvyswstrvgadlnmedllgssvypkryflintgtdlpilksneamvgalnmimgms 251

QY 252 LETVAPPNSKLERFTYHEHLERRVYEVKLPITNISKEAPPHNIOFVGSAYFVLISQAF 311

DB 252 mesevppkhketrwkyhfevvr---dlhl---tnkkkdpypnltmfignaylvasrdt 305

QY 312 VKIIFNNISVODFFRWSKDTYSPDEHFMATLIRVPGIGEL-SRSADVDSDLOSKTRLYK 370

DB 306 vqhwtkmpksqqllewxrdylspdehfwatqlqarmpgsvpnkylsdmsctatlvk 365

QY 371 WNYEGFP-----YPSCGTGSHLRVCITYGAELRWLTKDGHWFANKEDSKVDPIILIKCLA 425

DB 366 wqhgagdldkgapyapcsghlqrafcyvgagdlmwlmqnhlllankfdpkrvdnalgcle 425

QY 426 EKL 428

DB 426 eyl 428

RESULT 5

AAAG75058

ID AAG75058 standard; Protein: 465 AA.

XX

XX AAG75058;

XX

DT 03-SEP-2001 (first entry)

XX

DE Human colon cancer antigen protein SEQ ID NO:5822.

XX

XX Human: colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 15.

XX

OS Homo sapiens.

XX

XX WO200122920-A2.

PD 05-APR-2001.

XX

XX 28-SEP-2000; 2000WO-US26524.

PE

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

PI

XX WPI; 2001-235357/24.

DR N-PSDB; AAH34463.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers -

PT

XX

PS Claim 11; Page 7338-7340; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.

CC Additionally, N may be used to produce the colon cancer-associated P,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX

SQ Sequence 465 AA;

Query Match 36.3%; Score 868; DB 22; Length 465;

KW UDP-GlcNAc:Galbeta1-3GalNAc-alphaR; transgenic animal; germ line;
 KM beta-1,6-N-acetylglucosaminyl-transferase.
 KN
 OS Rattus sp.
 OX
 PN CA2186987-A.
 PP
 PD 02-APR-1998.
 PX
 PF 02-OCT-1996; 96CA-2186987.
 PG
 PR 02-OCT-1996; 96CA-2186987.
 PS (MOUN) MOUNT SINAI HOSPITAL CORP.
 PT Dennis JW, King GL, Koya D, Nishio Y, Warren CE;
 PU WPI: 1998-339608/35.
 PV N-PSDB; AAX24042.
 PW
 PY Screening for substances that prevent or treat cardiomyopathy
 PZ associated with diabetes and hyperglycaemia - compises reacting
 QA core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
 QB in presence of test substance
 QC
 QD
 QE
 QF
 QG
 QH
 QI
 QJ
 QK
 QL
 QM
 QN
 QO
 QP
 QQ
 QR
 QS
 QT
 QU
 QV
 QW
 QX
 QY
 QZ
 RA
 RB
 RC
 RD
 RE
 RF
 RG
 RH
 RI
 RJ
 RK
 RL
 RM
 RN
 RO
 RP
 RQ
 RS
 RT
 RU
 RV
 RW
 RX
 RY
 RZ
 SA
 SB
 SC
 SD
 SE
 SF
 SG
 SH
 SI
 SJ
 SK
 SL
 SM
 SN
 SO
 SP
 SQ
 SR
 SS
 ST
 SU
 SV
 SW
 SX
 SY
 SZ
 TA
 TB
 TC
 TD
 TE
 TF
 TG
 TH
 TI
 TJ
 TK
 TL
 TM
 TN
 TO
 TP
 TQ
 TR
 TS
 TT
 TU
 TV
 TW
 TX
 TY
 TZ
 UA
 UB
 UC
 UD
 UE
 UF
 UG
 UH
 UI
 UJ
 UK
 UL
 UM
 UN
 UO
 UP
 UQ
 UR
 US
 UT
 UV
 UW
 UX
 UY
 UZ
 VA
 VB
 VC
 VD
 VE
 VF
 VG
 VH
 VI
 VJ
 VK
 VL
 VM
 VN
 VO
 VP
 VQ
 VR
 VS
 VT
 VU
 VV
 VW
 VX
 VY
 VZ
 WA
 WB
 WC
 WD
 WE
 WF
 WG
 WH
 WI
 WJ
 WK
 WL
 WM
 WN
 WO
 WP
 WQ
 WR
 WS
 WT
 WU
 WV
 WW
 WX
 WY
 WZ
 XA
 XB
 XC
 XD
 XE
 XF
 XG
 XH
 XI
 XJ
 XK
 XL
 XM
 XN
 XO
 XP
 XQ
 XR
 XS
 XT
 XU
 XV
 VX
 VY
 VZ
 YA
 YB
 YC
 YD
 YE
 YF
 YG
 YH
 YI
 YJ
 YK
 YL
 YM
 YN
 YO
 YP
 YQ
 YR
 YS
 YT
 YU
 YV
 YW
 YX
 YY
 YZ
 ZA
 ZB
 ZC
 ZD
 ZE
 ZF
 ZG
 ZH
 ZI
 ZJ
 ZK
 ZL
 ZM
 ZN
 ZO
 ZP
 ZQ
 ZR
 ZS
 ZT
 ZU
 ZV
 ZW
 ZX
 ZY
 ZZ

[illegible]

QY 305 FVLSQAFVKIKIFNNISVIODFPMASDKDTSPPDHFNATILIRPGIRGELSRSAO-DVSDLD 363
 Db 289 fvvrelyevyvlcnkhidqfitemaqdltvspdeifwtatqiripvepgsfpsskhydlsdm 348
 QY 364 SKRFLVWNNYEGGF-----YPSCTGSHLRVCIIGAAELMLLIKDHMFANRDKSVDP 418
 Db 349 avarfhwkwyfeggdvngapypgpcsgvhrsvcvfvgdglswmlrkhhffankfdmdvdp 408
 QY 419 ILITCLAEKLEED 431
 Db 409 falqcleehlrhk 421

RESULT 8
 AAB54344
 ID AAB54344 standard; Protein; 465 AA.
 AC AAB54344;
 XX 09-MAR-2001 (first entry)
 DT
 XX Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
 DE
 KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.
 XX
 OS Homo sapiens.
 XX
 PN MO200055320-A1.
 PD 21-SEP-2000.
 XX
 PE 08-MAR-2000; 2000MO-U0505989.
 PR 12-MAR-1999; 9905-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM;
 DR WPI: 2000-579444/54.
 DR N-PSDB: AAC99109.
 XX
 XX New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 PS Claim 11; page 1237-1239; 1379pp; English.
 XX
 CC AAC98773 to AAC9921 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54408 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The

RESULT 8
ID AAB54344 standard; Protein; 465 AA.
XX
AC AAB54344;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:796.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytosolic; neuroprotective;
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
PN NC200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-579444/54.
XX
DR N-PSDB: AAC99109.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 11; Page 1237-1239; 13796P; English.
XX
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiac and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The

CC proteins can be used to treat or prevent neural, immune system, muscular, CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or CC proliferative disorders. AAC9232 to AAC9240 and AAB54467 represent CC sequences used in the exemplification of the present invention.

XX Sequence 465 AA;

Query Match 36.2%; Score 866; DB 21; Length 465;
Best Local Similarity 43.0%; Pred. No. 9.3e-70;
Matches 182; Conservative 74; Mismatches 143; Indels 24; Gaps 10;

QY 23 LMLLS--LKLINVRRLFPQK---DIYLVESLSTSPFVFN-RYTHVKDFEYEVNCSG 75
DB 40 lwalqymllatvalklisfxikcdedhlglesresgqyrnllynfiklpakrslncsg 99
QY 76 IY-FOEDL--EIGKSLIRRRDIDLEDVVAMTSQCDIYQTLRGIAQKLVKESKF 131
DB 100 vrrgqdeavlgallnlnlvkkr-efldthylsltdcehikaeakflgfpiskveef 158
QY 132 PLVSLVYHKDAIWERLIIHAIYNOHNYIC-IHYDRKAPDFKVMNNLAKCFNSIFIAK 191
DB 139 playamvthekienterilravaypqrlvcvhdekspeetkeavkaliscfpnvflask 218
QY 192 LEAVYAHISRQADNLNLSLSSIQWKVYINLCQDFPLKSNFELVSEKLNAGNM 251
DB 219 lrvvyaaswsvrqadlncmedllgsyvpwkyflncgtdfipksnaemvqalnmhngns 278
QY 252 LETVPPSKLERFYHHELRVPEYVKLPIRTNISKAPPHNIQIFVSGAVYLSQAF 311
DB 279 mesevpphketrvkwyhevr---dthl---cnkkdpppyllmtmgtaiyvaasrd 332
QY 312 VAITNNSTVQDFEAMSKDTISPDHEFWATLIRVPGIGEL-SSAQDVSDLSQKTRLVK 370
DB 333 vghvlnkpksgqllewvdclyspdehltqratmpgsvpnbpkydismtslarlvk 392
QY 371 WMYVGGFF----YPSCTGSHLRVCIGAAELRWLIKDGHWMPKRFSDKVDPLIKCLA 425
DB 393 wghgagddidkgapapcsqilqratvcyagadlnwmilgnhllankfpkvdnaicgle 452
QY 426 EKL 428
DB 453 eyl 455

RESULT 9
ID AAM93943 standard; Protein: 428 AA.
AC AAM93943;
XX
XX
DT 28-JUN-1999 (first entry)
DE Human core 2 GNT protein.
XX
XX
KW Screening; treatment; prevention; cardiomyopathy; inhibitor;
KW diabetes mellitus; hyperglycaemia; core 2 GlcNAc-T; acceptor substrate;
KW UDP-GlcNAc:Galbetail-3GalNAc-alpha1ar; transgenic animal; germ line;
KW beta-1,6-N-acetylglucosaminyl-1-transferase; human; core 2 GNT.
XX
OS Homo sapiens.
XX
XX CA2186867-A.
PN
XX
XX 02-APR-1998.
PD
XX
XX 02-OCT-1996; 96CA-2186987.
PF
XX
XX 02-OCT-1996; 96CA-2186987.
PR
XX
XX (MOUN) MOUNT SINAI HOSPITAL CORP.
PA
XX
XX Dennis JW, King GL, Koya D, Nishio Y, Warren CE;

XX WPI: 1998-399608/35.
DR N-PSDB; AAX24043.

PT Screening for substances that prevent or treat cardiomyopathy
PT associated with diabetes and hyperglycaemia - complees reacting
PT core 2 GlcNAc-T with acceptor substrate and sugar nucleotide donor
PT in presence of test substance
XX

PS Disclosure; Fig 9; 35pp; English.

CC This invention describes a method for screening for a substance that
CC may be used to prevent or treat cardiomyopathy associated with diabetes
CC and hyperglycaemia. This method involves reacting core 2 GlcNAc-T
CC (UDP-GlcNAc:Galbetail-3GalNAc-beta-1,6-N-acetylglucosaminyl-
CC transferase) with an acceptor substrate and a sugar nucleotide donor in
CC the presence of a test substance under conditions whereby the core 2
CC GlcNAc-T produces a reaction product, determining the amount of reaction
CC product, and comparing the amount of reaction product with the amount
CC obtained in the absence of the test substance, where lower amounts of
CC reaction product in the presence of the test substance indicate that the
CC substance inhibits core 2 GlcNAc-T. The invention also describes (1)
CC methods for preventing or treating cardiomyopathy associated with
CC diabetes and hyperglycaemia, and (2) transgenic nonhuman animal whose
CC germ cells and somatic cells all contain a DNA construct introduced into
CC the animal or an ancestor of the animal at an embryonic stage, where
CC incorporation of the DNA construct into the germ line of the animal
CC causes the animal to develop cardiomyopathy similar to that associated
CC with diabetes mellitus and hyperglycaemia. This sequence represents the
CC human core 2 GNT protein (beta-1,6-N-acetylglucosaminyltransferase)
CC which is used to describe the method of the invention.

XX Sequence 428 AA;

Query Match 36.1%; Score 863.5; DB 19; Length 428;
Best Local Similarity 43.1%; Pred. No. 1.4e-69;
Matches 187; Conservative 72; Mismatches 132; Indels 43; Gaps 12;

QY 28 LKLINVRRLF--PQKDIYLV-EYSLSTSPFVNRRTHWKDE--VRY-----EV 71
DB 1 mirltllrrllfsytkyfmvlslstfslr---lhqkpefsvrvrhlagempsd 57
QY 72 NCSGIYFOEPLIEIGK-----SLFIRRRDIDLEDVVAMTSQCDIYQTLRGIAQKLV 125
DB 58 nckvlgdvneiqkvkllvfkkrp--rtvpdyimscdcsfiktiryiveps 115
QY 126 KEKSPFIAYSLVHKDAIWERLIIHAIYNOHNYIC-IHYDRKAPDFKVMNNLAKCFNS 185
DB 116 keeaeplayslvvhklemldrlraiympqrfycvhdcksedeyilaavmjiaescfen 175
QY 186 IFIAKLEAVEYAHISRQADNLNLSLSSIQWKVYINLCQDFPLKSNFELVSEKLN 245
DB 176 yvaasriesvvaswsvrqadlncmedllgsyvpwkyflncgtdfipksnaemvqalnmhngns 235
QY 246 LGAANMLETVKPNKSLERFYHHELRVPEYV--KLPIRTNISKAPPHNIQIFVSGA 303
DB 236 lmgemlntetrmshkeery-----kryevngkl-tntgvtvmlpleeplsgsa 287
QY 304 YFVISAQAFYIIRNNSTVQDFEAMSKDTISPDHEFWATLIRVPGIGELISRSNO-DVSD 362
DB 288 yfvvsvreyvayvlgneqiklmewagdclyspdehltqratmpgsvpnbpkydismtslarlvk 347
QY 363 QSKTRLVKMYVYGGFF----YPSCTGSHLRVCIGAAELRWLIKDGHWMPKRFSDKVD 417
DB 348 gavarfvkwgyfgdvskgapypcdgghvsvclfgagdlmvlrkhllfankfvdvd 407
QY 418 PILIKCLAEKLEBO 431
DB 408 lfaigcdelnrhk 421

RESULT 10

Query Match	36.1%;	Score 862;	DB 21;	Length 406;
Best Local Similarity	44.5%;	Pred. NO. 1.8e-69;		
Matches 173; Conservative	70;	Mismatches 128;	Indels 18;	Gaps 8

```

QY 51 STSPFVNN-RTHVHKDEHYEVNNSGTY--EOEP--EIKSLEIRRDIDDEDVVA 105
D 15 sqsgycenlllynfiklpakrsincsgytrgdqevlqalnnlevvkkr-epftdchyls 73
QY 106 MTSODDIYQTLRGYAQKIVSKVSEKSPRIANSLVYHKDAIWERLIIAIYMOHNIYCIHYD 155
D 74 ltrccenikaerkfifqfolskeevelfaysmvlnheklenferllravyarqnlcyhvnd 133
QY 166 RKADDTKRVANNNLAKCFNSNFIASKILEANEYAHITSLOADLNCISDLSSIQMKVYN 225
D 134 ekspektieaakhaliscprvfiasklvrvyaaaswsvqgdlncomedllqssvwpkylfn 193
QY 226 LCGDPPFKSNFVELSELKTLGNANMLETYKPPRSKLERFTYHHELRYRVEYUKLPIRT 285
D 194 tccgdcfrfksnaemqaklmngnsmesevppkhketrwkyhneivr---dclhl--t 247
QY 286 NISKEAPPHNTQIFVGSAYFYLSQAFKYLIFNNSIVODFPAMSKDYISPDHEFWALIRY 345
D 248 nkkddpprynltnftgmaylvassdfvghvlnkpxsqgllwexvdtyspdehlatqlqra 307
QY 346 PGIQGET-SRBAQDVSDIQSKTRLVKKNMYDEGF----YPSCTGSLRSVCYIGAALR 399
D 308 twmpgsyvpnhpkysldamtstariwkvqghiegdldkqayapcsqjngraqrcvysgaqdn 367
QY 400 WLKGDHMFANKPDSKVDPLIKLAKEL 428
D 368 wmlqnhhllankfcpkvcdnaalqclceely 396

```

RESULT	12
AAR51386	
ID	AAR51386 standard; Protein; 428 AA.

XX	AAK51386;	
XX		
AC		
DT	07-OCT-1994	(first entry)
XX		
DE	Sequence of human core 2 beta 1-6	
DE	N-acetylglucosaminyltransferase (C2GnT or core 1-6 Agt).	
XX		
KW	C2GnT: 1-6 AGT; core 2 beta 1-6 n-acetylglucosaminyltransferase	
KW	O-glycan.	
XX		
OS	Homo sapiens.	
XX		
XX		
XX		
FT	Domain	Location/Qualifiers
FT		10..32
FT	Modified-site	/label= signal/membrane anchoring domain
FT		52
FT		/label= potential N-glycosylation site
FT		/note= "see also Aas 58 and 95"

```

PN      EP590747-A.
XX
XX
PD      06-APR-1994.
XX
XX
PF      29-SEP-1993;      93EP-0250268.
XX
XX      01-OCT-1992;      92US-0955041.
PR
XX
XX      (LJOL-) LA JOLLA CANCER RES FOUND
PA
XX
XX      Biehuizen MFA, Fukuda M;
PI
XX
XX      WPT, 1994-111195/14.
DR      N-PSDB; AA061559.
XX
XX

```

PT New beta 1-6-N-acetylglucosaminyl transferase and acceptor-
PT used for the study of the effect of variant O-glycan(s) on
PT cell-cell interactions, partic. in cancers
XX
XX Disclosure, Page 20-22; 34pp; English.

CC C2GnT, or an active fragment thereof, catalyses the formation of
CC critical branches in O-glycans. cDNAs encoding various
CC glycosyltransferases can be isolated by transient expression of cDNA
CC in recipient cells, e.g., Cos-1 cells were transfected
CC with a cDNA library, pcDR alpha-2F1, constructed from poly(A)⁺ RNA
CC of activated T lymphocytes which express the C2GnT. Transfected
CC cells were selected using MAb T305, which identifies a
CC hexasaccharide on leukostallin. Leukostallin CD33 is an acceptor
CC molecule for C2GnT activity. A plasmid, pcDR alpha-Ten,
CC which directed expression of the T305 antigen was identified. The
CC cDNA insert was isolated and sequenced. The cDNA encoded the
CC acceptor molecule leukostallin CD33.

SQ Sequence 428 AA;

Query Match	35.88;	Score 854.5;	DB 15;	Length 428;
Best Local Similarity	42.98;	Pred. No. 9e-69;		
Matches 186;	Conservative 72;	Mismatches 133;	Indels 43;	Gaps 12;

```

QY 28 ILKILNRLRF--PQKDIYLV-EYSLSTSPVNRNRYTHVQDE---VVR-----EV 71
Db 1 mlrlrlrrllfslpckkyyfmlvlslstlffsvlr---lbgkpefsvrhlajlagenpsadi 57
QY 72 NCSGIEODEPLEIGK-----SLETRRDITLEDDEDDVAMTSQDITQYLRGAOKLVs 125
Db 58 nctkylgdcvneiqvkllelltvkfkkrp--rtwpeddylnmtsdossfikrpkylvlppls 115
QY 126 KEENSEFPLAYSIVYKKDALMYERLCHAIYNOHNICYCHYDKKADDTFVANNNAKCESN 189
Db 116 Keeneefplaysivvnhklemldrlsalylnqnlvcvhwddksdsdasyaaamgascfn 175
QY 186 IFAIAKLEAVEYAHISRLQADLNCLSDLKSSIQWKYVINDGQDFPLKSNFELVSEIKR 245
Db 176 vlvastlesvvyasvrsrgadlnemckldylamsfmkyljlnlgmddfprktnlelvvrkl 235
QY 246 LNGANMLETVPKPNKSLERFYYHNEELRRVRYEYV--KLPRTNISKENAPRNINQIPLVQSA 307
Db 236 lmgemlletempshkeerV-----kkrlyevangkl-tlrgtvmklprletpdlfsgsa 287
QY 304 YFVLQAFKRYTFENNSIYQDFEAFMKDQYSPDEHMAWLLRPVGGIPGRISASQ-DVSDL 367
Db 288 ylvvrerayvgvylqnekqlkklmeweqddyspdeylwaclqrlpvepgslpashkydyldsm 347
QY 363 QSKTRLVWKNYNYEGFF-----YPSCTGSHLSVCITCYGAELRWLlKGDHWPANKFDSKYD 411
Db 348 qavarfvfwqyfegdvsvskgapyrpecdyghvsvclfsggdjlnmlrlkhhllfankfdvdvd 407
QY 418 PLIKICLAKLEEQ 431
Db 408 llaigcldehlnhk 421

```

XX	RESULT 13
XX	AAB30297
XX	ID AAB30297 standard; Protein; 427 AA.
XX	
AC	AAB30297;
XX	
XX	12-FEB-2001 (first entry)
DE	
XX	Diabetic rat heart core 2 GlcNAc-T.
XX	
XX	
KM	Rat; diabetes; core 2 GlcNAc-T; cardiomyopathy; hyperglycaemia
KM	UDP-GlcNAc:galbeta1-3galNAc:alphaGal beta1-6
XX	N-acetylglucosaminyltransferase.
XX	

OS Ratus sp.
 XX US6131578-A.
 PN 17-OCT-2000.
 PD 02-OCT-1997; 97US-0943058.
 PE 02-OCT-1996; 96US-0046876.
 PF (KING/) KING G L.
 PG (NISH/) NISHIO Y.
 PA (KOYA/) KOYA D.
 PB (DENN/) DENNIS J W.
 PC (WARR/) WARREN C E.
 PD Nishio Y, Koya D, King GL, Warren CE, Dennis JW;
 DR WPI, 2000-678642/66.
 DR N-PSDB; AAC65468.
 XX Preventing or treating cardiomyopathy associated with diabetes mellitus
 PT and hyperglycemia, comprises administering a substance that inhibits
 PT core 2 N-acetylglucosaminyltransferase activity -
 PS Example 1; Fig 3b; 21pp; English.
 XX The present invention describes the human UDP-GlcNAc:Galbeta1-3GalNAc
 CC alpha6 beta1-6 N-acetylglucosaminyltransferase (core 2 GlcNAc-T) protein
 CC and coding sequences. The enzyme is associated with cardiomyopathy in
 CC diabetes and hyperglycaemia sufferers. The invention also provides
 CC methods for identifying inhibitors of core 2 GlcNAc-T which can be used
 CC in treatment.
 CC XX Sequence 427 AA:
 SQ
 Query Match 35.6%; Score 850; DB 21; Length 427;
 Best Local Similarity 40.9%; Pred. No. 2.3e-68;
 Matches 118; Conservative 78; Mismatches 133; Indels 46; Gaps 12;
 QY 28 LKLLVRLRLE--PQRDIYV-EYSLSTSPFVNR-----YTHVK--DEVREVNCS 74
 DB 1 mlrlrrrrlfsyptkyfmlvvlslflevrhhqkpefsvshlelbgddpsnvnct 60
 QY 75 GIVEOEPLIEGKSLERBRDIDLEP-----DDVAMTSGCDIYQTLRGYACKIVSK 126
 DB 61 kvlgqdeekqkv---kklrtvgekkrrptphyeitmtctdcaeflrrkyimeplk 116
 QY 127 EKSFPRLASLVYHKDAIWERLHAIFYNQHNITCIHYDRKAPDFEFKAMNNIAKCFNSI 186
 DB 117 eevgfplasyivhkhldmrlrralymprnycihvdrkaeesflaavqgiasefdm 176
 QY 187 FISKIEAYEAHISRLDNLGSLDLSKSIOMKVVINLCODPFLKSNFELVSELKL 246
 DB 177 lvasqlesvayvaswvkdnlcmkdlymnaakylldlqmdlplknlhivtklflsf 236
 QY 247 NGANMLETVK-PPNSKLERITYHHELRVRYEV--KLPRTNISKAPENITQIEVGS 302
 DB 237 tgsnletempnke-----kvwkyryvvdgkl-tltygvkaaprlklplffgs 285
 QY 303 AYFVLSQAFVKYIFNNSIYODFFAWSKDVTSPDEHFMALILINPGISGSSAQ-DVSD 361
 DB 286 ayfvvleayvgylenkngkltmewagdtyspelfatqltpevgslpsshkyldsd 345
 QY 362 LQSKTRLVKNNYEGF-----YPSCTGSHLSRVCITGAAELRLIKDGHMFAKFDISKV 416
 DB 346 mnavarfvkqglfegdvsgagpyppsgyhrsvcvfygdlswmlrkhhfankfcmv 405
 QY 417 DPLIKCLAEKLEEQ 431
 DB 406 dplalqcleehlrhk 420

RESULT 14
 AAR71932
 ID AAR71932 standard; Protein; 400 AA.
 XX
 AC AAR71932;
 XX
 DT 23-SEP-1995 (first entry)
 XX
 DE I-branching enzyme.
 XX
 KW Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IgG;
 KW inflammation; antiinflammatory; tumor; hypersensitivity; anemia;
 KW transgenic animal; I antigen.
 XX
 CS Homo sapiens.
 XX
 FN W09507020-A.
 XX
 PD 16-MAR-1995.
 XX
 PF 09-SEP-1993; 93WO-US08476.
 XX
 PR 09-SEP-1993; 93WO-US08476.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Bierhuizen MFA, Fukuda M;
 XX
 DR WPI: 1995-123182/16.
 DR N-PSDB; AA089201.
 XX
 PT Nucleic acid encoding I-branching enzyme - used to develop prods.
 PT for treating e.g. inflammatory responses, tumours or
 PT hypersensitivity reactions
 PS
 PS Disclosure; Fig. 6; 56pp; English.
 XX
 CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was
 CC reverse-transcribed and inserted into expression vector pCDNA1.
 CC Plasmid DNA was used to transfect CHO-Py-leu cells. Transfected
 CC cells were screened using human anti-I antigen antibodies and goat
 CC anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in
 CC AA089201) encoding I-branching enzyme (AAR71932).
 CC XX Sequence 400 AA:
 SQ
 Query Match 29.8%; Score 711; DB 16; Length 400;
 Best Local Similarity 43.7%; Pred. No. 7.6e-56;
 Matches 143; Conservative 43; Mismatches 129; Indels 12; Gaps 5;
 QY 108 SDCDIYOTLRGKIVACKIVSEKSPFIASLVYHKDAIWERLHAIFYNQHNITCIHYDRK 167
 DB 70 ssckeylqshytlapiskaeadflaylmvhhftfarlfralymprnycihvdrka 129
 QY 168 APDTFKVAMNNIAKCFNSIFASKLEAVEYAHISRLDNLGSLDLSKSIOMKVVINLC 227
 DB 130 atfekdavegljascfnafaskmepvvgysrlqadlnctrdlsafeswkyvntc 189
 QY 228 GDDPFLKSNFELVSELKLNANMLETVKPPNSKLERITY-HHELRVRYEYKLPRTN 286
 DB 190 qddflpknkeltvylkqfkgknltpvlpahalgtrkyvhghlsgkelsyv--lrlt 246
 QY 287 ISKEAPENITQIEVGSAYFVLSQAFVKYIFNNSIYODFFAWSKDVTSPDEHFMALILIRP 346
 DB 247 alkppppnllilygsayvalsrfeanfvlhdpravdllyskdtkfspdehfwlrlrip 306
 QY 347 GIPGEISRSADVSDQSKTRLVKNNYEGF--YPSCTGSHLSRVCITGAAELRLINDGH 406
 DB 307 gvpssmpnas-----wtgnlraikwsdmed-rhggchghyvhgiclyngldkwlvsps 360
 QY 407 WFANKFDSKVPDILIKCLAEKLEEQR 433

Page 1

617 ++++++CCCCC+C+C+a+r+C
0v


```

Db 970 acctcaactgcacggaagactctccacagctcagtcagtcgctggaataacttccgaata 1029
Qy 677 tctgtggcgaagatttcccccgaagtaatttgaattgtgacagagtcgaataaac 736
Db 1030 cctgtggcgaagacttccctataaagacagatgacagatgctcagctcgaagatg 1069
Qy 737 tcaatggagcaaatatgttggagcagtgtaaaccccccaacagtaatttgaagaatca 796
Db 1090 tgaatggagcagatgacagatgacagatgacagatgacagatgacagatgacagatg 1149
Qy 797 ctaccatcatgaacttgaagcagtgctcctatgaatatgtgaagctaccataagaaac 856
Db 1150 aatcatcattgaggaagtgagagacacattacac-----ctaacca 1191
Qy 857 aatctccagaagaaacaccccccaataacatcagataattgtgacagtcatttg 916
Db 1192 acaagaagaagatccctcccttataatttaactatgtttacaggaatgctgacatg 1251
Qy 917 ttttaagtcagcatttcttaatatatttttaacaactcctgcttaagaacttttg 976
Db 1252 tgggtcccgagatttcgacacatgttttgaagacccctaaatcccaaacatgattg 1311
Qy 977 cctgtgtcaaaagacacatcctcctgtagagacatttggctacattgctgagctc 1036
Db 1312 aatgggtaaagacacattatagccagatgacacacctctggccaccccttcagcgtgac 1371
Qy 1037 caggaataccctggagagatttccagatcagcccaagatg---tgctatctgcagagta 1093
Db 1372 ggtgtagctgtgctgtctgtcccaacaccccaagatgacatcccaacatgacttcta 1431
Qy 1094 agactgcctgtcaagtgagtaactatgagagcttttctcactccagct----- 1143
Db 1432 ttgcagagctgttcaagtgagtgagtgatgagtgagtgagtgagtgagtgagtgagtg 1491
Qy 1144 ----tgatctgagctcactcctcgaaagcgtgtatattatgagtgagctcagaatgaagct 1198
Db 1492 ctccctgctgtgacccacccagcgggtatctgctgtatgagtgagtgagtgagtgagtg 1551
Qy 1199 gacttcaaaagatgagcattgttgcataaatttgcataaatttgcataaatttgcataa 1258
Db 1552 ggaatgctcaaaacacacacacacacacacacacacacacacacacacacacacacac 1611
Qy 1259 tgatgaatgctgtgcagagaagct 1283
Db 1612 cctctcagtgcttagaagaataact 1636

RESULT 2
US-08-118-906-13
; Sequence 13, Application US/08118906
; Patent No. 5484590
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Biehuizen, Marti F.A.
; TITLE OF INVENTION: Expression of the Developmental I
; TITLE OF INVENTION: Actigen By a Cloned Human cDNA Encoding a Member of a
; TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
; NUMBER OF SEQUENCE: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,906
; FILING DATE: 09-SEP-1993

```

```

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,615
; REFERENCE/DOCKET NUMBER: P-LI 9526
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 255..1454
; US-08-118-906-13

```

```

Query Match 13.6%; Score 185.2; DB 1; Length 1807;
Best Local Similarity 52.1%; Pred. No. 7.2e-43;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

Qy 327 ttgtgacattatcagctcctaaagctatgctcaaaagctgtctcgaagagagaga 386
Db 467 TTGCAAGGAATACTTGACCCAGAGCCACTACATCACGCCCTTTATCTAAGGAAGAAC 526
Qy 387 aagctcccaagacacacacacacacacacacacacacacacacacacacacacacacac 446
Db 527 TGACTTCCCTGGCATATATATATGTCATCATCATCATCATCATCATCATCATCATCATCAT 586
Qy 447 tatccatgcatataacacacacacacacacacacacacacacacacacacacacacacac 506
Db 587 CTTGAGGGCATTTTACATGCCCCCAAAATATCTACTGTTCATGATGATGATGATGATGATG 646
Qy 507 tgatactcaagctgacatgacatgacatgacatgacatgacatgacatgacatgacatgac 566
Db 647 AATGATTTTAAGATGCGGTAGAGCAATATTAAGCTGCTCCAAAGCCTTTTGTGGC 706
Qy 447 tatccatgcatataacacacacacacacacacacacacacacacacacacacacacacac 506
Db 587 CTTGAGGGCATTTTACATGCCCCCAAAATATCTACTGTTCATGATGATGATGATGATGATG 646
Qy 567 ttcaaatgagagctgtggaatatgccacattccagacattccagagctgaattaatg 626
Db 707 TTCCAAAGTGAACCCCTGTCTATGAGGAGATCTCCAGGCTCAGGCTCAGGCTCAGGCTC 766
Qy 627 ctgtgacacctcgaagctcctcaatccagtggaatatgtatacaactgtgtgga 686
Db 767 CATCAGAGATCTTCTGCTTCCAGGCTCATGGAAGTATGATATGATATGATATGATATGAT 826
Qy 687 agatttccctgaagcacaatttgaattggtgctcagagtgtaaaaaactcaatgagc 746
Db 827 AGACTTCCCTGTAAGCAACAGAAATAGTCACTATCTAAGAGATTTTAAGCTAA 886
Qy 747 aatcatgttgagagcgttgaaaccccccaacagtaatttgaagaatttcaactacatca 806
Db 887 AAATATCAGCCAGAGGCTGCTGCCCCAGCTCATGCAATTTGAAGAGCTAAATATGCTCA 946
Qy 807 tgaactagacggtgtccttatgatatgtgaagctacccaataagacaacatcccaac 866
Db 947 CCAAGAGACACTTGCGCA-----AAGAGCTTCTCTATGATATGAAGCAACAGCTTGA 1000
Qy 867 ggaagacaccccccaataacatcagataatttggagagtgcttaattttaaagta 926
Db 1001 ACCGCTCCGCCCATATATCATCAAAATTTACTTTGGCTGCTCATATGCTCATATCAG 1060
Qy 927 agcatlgttaataatatttcaacaacatccatcgttcaagacttttgcctggtctaa 986
Db 1061 AGAGTTTGCCAACTTGTCTGCTGATGACCAAGGCGGTGTGATTTGCTCCAGTGCTCAA 1120
Qy 987 agacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1046
Db 1121 GGAACATTTCACTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1180

```

QY 1047 tgggagatttccagatcaagccagagatgtctatctgcagtaagtaactgcctgt 1106
 Db 1181 TGGCTCTATGCCAATGATCCTGACAG-----GAAACTCAGAGCTAT 1225
 QY 1107 caagtgaattactagaaggttttctatccagttgtaactgtcccttcagag 1166
 Db 1226 AAGAGGAGGTGACATGGAAGACAGACAGAGAGC---TGCCAGGCGCACTATGTAATG 1282
 QY 1167 cgtgtgatttctgagactgcagaattaggtgtctatcaagaatgacattgttc 1226
 Db 1283 TATTGTATCTATGGAAGAGGAGACTTAAGTGTCTGTTAATTCACCAAGCTGTTC 1342
 QY 1227 taataattgattcttaaggtgagacctctgtcttaagctgtgagaaagcttga 1286
 Db 1343 TAAACAAGTTGAGCTTAATACCTACCCCTTACTGTGAAATGCTAGAACTGAGGCAATCG 1402
 QY 1287 agaa 1290
 Db 1403 CGAA 1406

RESULT 3

US-08-486-196-13
 ; Sequence 13, Application US/08486196
 ; Patent No. 5731420

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
 APPLICANT: Bierhuizen, Marti F.A.
 TITLE OF INVENTION: Expression of the Developmental I
 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESS: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/486,196

FILING DATE:

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118, 906

FILING DATE:

ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER:

P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO:

13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

MOLECULE TYPE:

CDNA
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-486-196-13

Query Match 13.6%; Score 185.2; DB 1; Length 1807;
 Best Local Similarity 52.1%; Pred. No. 7.2e-43;

Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

QY 327 ttgagacattatcaagcttaagaggttatgtcacaagctgtctcaaggaagaa 386
 Db 467 TTGCAAGGATACTGTACCCAGAGCCACTATCATCAGGCCCTTTATCTTAAGGAAGAAC 526
 QY 387 aagcttcccaatagacctattcttggttgtccacaagaatgcaattatggttaagact 446
 Db 527 TGACTTTCCCTTGCCATATATTAATGTCATCCATCACTACCTTTGACACCTTGCAAGGCT 586
 QY 447 tatccagcatatataagcaagcaaatattctgtatccattatcatatgatgtgaagacc 506
 Db 587 CTTAGAGGCTATTATTCATGCCCCCAAAATATCTACTGTTCATGATGATGATAAAGCAAC 646
 QY 507 tgataccctcaagttgcatatgaaccaattagtaagtgtcttcccaataatttcatcgc 566
 Db 647 AACTGAATTTAAGATGCGGTAGAGCACTATTAAAGTCTTCCCAAGCCTTTCTGGC 706
 QY 567 ttccaattgagagctgtgtgaatatatgccacatttccagaactcagagctgatttaattg 626
 Db 707 TTCCAAAGATGGAACCCGTTGTCTATGAGGAGATCTCCAGGCTCCAGGCTGACCTGAACGTG 766
 QY 627 ctgtcggacctctgaagcttcaatccagttgaaatatgttatcaactgtgtgtgaa 686
 Db 767 CATCAGAGATCTTCTGCTTCGAGGTCCTCATGGAAGTACGTATATCAACACCTGTGGCA 826
 QY 687 agatttccctgaagatcaaatatttgattgtgtcagaagtgaataaactcaatggagc 746
 Db 827 AGATTCTCCCTTGAAACCAACAGGAAATGATCTAGTATCTAAAGGATTTTAAAGTTAA 886
 QY 747 aaatgttggagacgttgaacccccaacagtaaatgtgaagattcaactaccata 806
 Db 887 AAATATACCCCAAGGGTGTCTGCCCAAGCTCATGCAATTGACGGAATTAATATGTCCA 946
 QY 807 tgaacttaagcgggtgtccttaataatgtgaagcttaacaaatgaagaaacaaactccaa 866
 Db 947 CCAAGAGCACTGTGGCA-----AAGAGCTTTCCTATGTATGAAGAAACAACGCTTGA 1000
 QY 867 ggaagaccccccaataacattcagatattgttgcagagcttatgttttaagta 926
 Db 1001 ACCGCCCTCCCCCATATATCTCAAAATTTACTTTGGCTGTGCTATGATGCTATCAAG 1060
 QY 927 agcaattgttaatatattttcaacaactccatcagttcaagactttttgctgtgtcaa 986
 Db 1061 AGATTGTGCCAATTTGTTCTGTGATGACCCAGGCTGTGATTCTCTCCAGTGTCCAA 1120
 QY 987 agaacactactctccatgaatgaagcaatttgggttaccttgattcgggttccagaatacc 1046
 Db 1121 GGAACCTTATAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1180
 QY 1047 tgggagatttccagatcaagccagagatgtctatctgcagtaagtaagtaactgcctgt 1106
 Db 1181 TGGCTCTATGCCAATGATCCTGACAG-----GAAACTCAGAGCTAT 1225
 QY 1107 caagtgaattactagaaggttttctatccagttgtaactgtcccttcagag 1166
 Db 1226 AAGAGGAGGTGACATGGAAGACAGACAGAGAGC---TGCCAGGCGCACTATGTAATG 1282
 QY 1167 cgtgtgatttctgagactgcagaattaggtgtctatcaagaatgacattgttc 1226
 Db 1283 TATTGTATCTATGGAAGAGGAGACTTAAGTGTCTGTTAATTCACCAAGCTGTTC 1342
 QY 1227 taataattgattcttaaggtgagacctctgtcttaagctgtgagaaagcttga 1286
 Db 1343 TAAACAAGTTGAGCTTAATACCTACCCCTTACTGTGAAATGCTAGAACTGAGGCAATCG 1402
 QY 1287 agaa 1290
 Db 1403 CGAA 1406

RESULT 4
 US-08-486-135-13

```

: Sequence 13, Application US/08488135
: Patent No. 5766910
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Minoru
: APPLICANT: Blerhuizen, Marti F.A.
: TITLE OF INVENTION: Expression of the Developmental I
: TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
: TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: #1.0, Version #1.25
: APPLICATION NUMBER: US/08/488,135
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/118,906
: FILING DATE: 09-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 9526
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1807 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 255..1454
: US-08-488-135-13

```

Query Match 13.6%; Score 185.2; DB 1; Length 1807;
 Best local similarity 52.1%; Pred. No. 7.2e-43;
 Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

QY 327 ttgtgacattatcagaacttaagagttatcctcaaaagcttgcctcaagaagagagaa 386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 ttgcaaggaatcttaccacccagacccacacacacacacacacacacacacacacac 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 aagcttcccaatagcctctcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 527 tgactttcccttgccatattatattgtctatccatctcttgcacacctttgcaaggt 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 tatcatgctatatatacaacacacacacacacacacacacacacacacacacacacac 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 587 ctctagggctattttacatcccccacaaatattctgtgtgtgtgtgtgtgtgtgtgtgt 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 507 tgtacctcaaggtgtgccttgagcaaataggtaagtgctctcacaataatttctatcgc 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 647 aactgaatttttaagatgcgttagacgaactatttgaactgttccaaagcctttttgtgc 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 567 ttccaaattagaggtgtgtgaatattcccaacattccacagctccagctgtatttaattg 626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 707 ttccaaattagaggtgtgtgaatattcccaacattccacagctccagctgtatttaattg 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 627 ctgtgcgacctcttgaggtcttcaatccagtgagaaatatgttatcaacttgtgtgagca 686

```

```

DB 767 catcagagatcttctgcccctgcaggtctcaggaatggaagttatcacaacacctgtgggca 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 agatttcccttaagtcacaatttgaattgtgtgaaggttgaataaacatcaatgagc 746
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 827 agacttccccccttaaaacacacacacacacacacacacacacacacacacacacacac 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 747 aaatagttgagagcgttgtaaccccccaacaglaatttgaagaagttacattaccatca 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 887 aaatattacccacaggggtgctgcccacagctcagcaatttgacgagcactaaatattgtcca 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 807 tgaacttaagacgggtgcttatgaataatgtaagtcacacacacacacacacacacacac 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 947 ccaagacacaccttgggca-----aagagcttttccttattgattagaacacacacgcttgaa 1000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 867 ggaagacaccccccaataacattcagataattgttgagagtgcttatttgtttaaagta 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1001 accgcccctcccccttaacacacacacacacacacacacacacacacacacacacac 1060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 927 agacttgttaataataatttcaacacacacacacacacacacacacacacacacacacac 986
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1061 agatttgcacacatttttttgcattgacacacacacacacacacacacacacacacacac 1120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 987 agacacactctctcgtatgagacatttggctacccttgatcgttggttccaggaatacc 1046
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1121 ggacaccttgcagtccttgcattgagcatttctgggtgacacacacacacacacacacac 1180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1047 tggagagatttccagatcagccagagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1181 tggcttatgccaatattcattcttgactg-----gaaactcagacactat 1225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1107 caatggaattactatgaagcttttctatcccaagttgactgacatcaccctgaag 1166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1226 aaagtgcagtgacatggaacacacacacacacacacacacacacacacacacacacac 1282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1167 cgtgtgtattatgagacacacacacacacacacacacacacacacacacacacacacac 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1283 tatttatttattggaacacacacacacacacacacacacacacacacacacacacacac 1342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1227 taataaatttgatcctgaagttgacccatctcttgatgaatgctgtgcagaaagcttga 1286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1343 taacaagttgagcttatttacttacccttactgtgtgaatgctgaactgagacatcg 1402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1287 agaa 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1403 cgaa 1406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5
 US-08-474-065-13
 Sequence 13, Application US/08474065
 Patent No. 5830465
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Minoru
 APPLICANT: Blerhuizen, Marti F.A.
 TITLE OF INVENTION: Expression of the Developmental I
 TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
 TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/474,065

FILING DATE: 424
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/118,906
 FILING DATE: 09-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1807 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 255..1454
 US-08-474-065-13

Query Match 13.6%; Score 185.2; DB 2; Length 1807;
 Best Local Similarity 52.1%; Pred. No. 7.2e-43;
 Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

QY 327 ttggacattatcagaactcctcaagaagttatgctcaaaagcttgctcacaaggagga 386
 DB 467 TTGCAAGGAACTCTTGACCCAGACCCACTCATCACGCCCTTACTAAGGAAGAGC 526
 QY 387 aagcttcccaatagcctattcttggtgtccacaagaatgacatttggttgaaggt 446
 DB 527 TGACTTCCCTTGCCATATATAATGTCATCCATCATCTTGTGACACCTTTGCAAGCT 586
 QY 447 tatccatgctataacaacagcagcaatattactgcatacatgatcgtgaagacc 506
 DB 587 CTTAGGGCTATTACATGCCCCAAATATCTACTGTCTGTTGATGATGATAAAGCAAC 646
 QY 507 tgataccctcaagctgacatgaacaattagcctgaagtgctctccacataatttcattgc 566
 DB 647 AACTGAATTTAAAGATCGCGTAGAGCAACTATTAAAGCTGTTCCCAACCTTTTCTGCG 706
 QY 567 ttccaattgagagctgtgaatatgccacattccagaactccaggtgatttaattg 626
 DB 707 TTCGAAGATGGAAACCGCTGTCTATGAGGGATCTCCAGGCTCCAGGCTGACCTGAAC 766
 QY 627 ctgtcggacctctgaagcttcaatccagtggaatatgcttatgaactgtgtggca 686
 DB 767 CATCAGAGATCTTCTGCTCCAGGCTCATGATGAGTACGTTATCAACCTGTGGCA 826
 QY 687 agatttccctgaagcaaatattgaattggtcagagtggaagaaactcaatgagc 746
 DB 827 AGACTTCCCTGAAACCAACAAGAAATAGTTCACTATCTGAAGATTAAAGATA 886
 QY 747 aatatgttgagagcgtgaagccccaacagtaaatggaaagatttcctacatca 806
 DB 887 AAATATACCCCAAGGAGTGTGCTGCCCAAGCTCATGCAATGAGAGCACTAAATATGCA 946
 QY 807 tgaactagaaggtgtgcttaataatgtgaagctaacataaaggaagaacacccca 866
 DB 947 CCAAGACACCTGGCA-----AAGAGCTTCTATGATGATAAGAACACACGTTGA 1000
 QY 867 ggaagcagcccccataaacttaagatattgttgcaagtgctatttgtttaaagta 926
 DB 1001 ACCGCCCTCCCCCATATATTCACAAATTACTTGGCTGTGCTATGTGCTTATCAAG 1060
 QY 927 agcattgttaataataatttcaacaactcactgcgtcaagacttttggcctggtctaa 986
 DB 1061 AGAGTTTGCCAACTTGTCTGATGATGACCCAGGCGTGTGATTGCTCCAGTGGTCAA 1120

QY 987 agacacatactctcctgatgagcaatttgggctacattgattcgggttccagaataacc 1046
 DB 1121 GGAACACTTTCAGTCCGATAGCATATTGCGTGACACATCAAGATTCCAGGTTC 1180
 QY 1047 tggggagatttccagatcagccagagatgtgtctgactcagagagaagactgcctgt 1106
 DB 1181 TGGCTTATGCCAAATGCAATCCTGGACTG-----GAAACCTCAGACCTAT 1225
 QY 1107 caagtggaattactgaagagcttttctatccagttgactggttccacttggaag 1166
 DB 1226 AAGTGAGTGCATGGAAGACAGACAGGAGC-----TGCCAGGCCACTATGTACATGC 1282
 QY 1167 cgtgtattctcagcgctgcagaatgaagtggtcttaacaagatgacattggttgc 1226
 DB 1283 TATTGTATCTATGGAACGAGACTTAAAGTGGCTGTTAATTACCAAGCCTGTTGC 1342
 QY 1227 taataattgattcctcaagtgagcctatcttgatataatgcttggcagaaagcttga 1286
 DB 1343 TAACAAGTTGAGCTTAATTAATCTACCCCTTACTGTGAAATGCTTAAGACTGAGCATGC 1402
 QY 1287 agaa 1290
 DB 1403 CGAA 1406

RESULT 6
 US-07-955-041-3
 Sequence 3, Application US/07955041
 Patent No. 5360733
 GENERAL INFORMATION:
 APPLICANT: FUKUDA, MINORU
 APPLICANT: BIERHOIZEN, MARTI FA
 TITLE OF INVENTION: A NOVEL BETA-6
 TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
 TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CAMPBELL AND FLORES
 STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
 CITY: SAN DIEGO
 STATE: CALIFORNIA
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/955,041
 FILING DATE: 19921001
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: CAMPBELL, CATHRYN
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9294
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-9001
 TELEFAX: 619-535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2105 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: both
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 220..1504
 FEATURE: polyA_signal
 NAME/KEY: polyA_signal
 LOCATION: 1913..1918


```

Db 549 TGTAAACCCCTTAAAGAGAGGCGAGTTTCCAAATACATATTTGATAGGTTCA 608
QY 420 caaagatgaatattatggtgaaaggcttatccatgctatatcaaacagccatattta 479
Db 609 TCACAAGATGTAATGCTTGACAGGCTCTGAGGCCATATATAGCCTCAGATTTTCTA 668
QY 480 ctgcatcattatgctgtaaggacccgtataaccttcaagttgccatgaaataggc 539
Db 669 TTGCGTTATGTTGACACAAAATCCGAGGATTTCTATTACTGACGATGAGTGGCCATCGC 728
QY 540 taagtgcttcccaattttccatgcttcccaattgaagctgtggaatagcccaat 599
Db 729 TTCCGTTTATGATATGCTTTTGCCAGCCGATGAGAGTGGTGTATGATGCTG 788
QY 600 ttccagatccaggtctgatttaattgcttgcgaaccttgcgaagcttcaatccagtg 659
Db 789 GAGCGGGGTGAGGCTGACCTCACTGATGAGGATCTCTATGCAATGAGTGCMAACTG 848
QY 660 gaaatgttatcaacttggtggaagattttcccttgaaagcaatttgatgtg 719
Db 849 GAAGTACTGTAATATCTTTGTGTATGATTTTCCATTAAACCACTAGAAATTTGT 908
QY 720 gtccagatgtaaaaaactcaatggaagcaatattgtggaagcggtgaaaccccaacag 779
Db 909 CAGGAAGCTCAAGTTGTTAATGCGAATAACACCTGGAACGAGAGATGCCATCCCA 968
QY 780 taatgtgaagatcattccatccatgaacttagaaggtgcttatgatatgaa 839
Db 969 TTAAGAAAGAGGTTGGAAGAGCGGTATGA-----GGTCTTAAATGGAAGCGTAC 1019
QY 840 gctaccaataaagacacatctccaaaggaagcaaccccccataacattcagatatgt 899
Db 1020 -----AAACACAGGAGACTGCAAAATCTTCCTCCATCGAACAACCTCTTTTC 1070
QY 900 tggcagtgctatttgttttaagtaagcatttgttaataatatttcaaacatccat 959
Db 1071 TGGCGTCCCTACTTCGTCGTCAGTAGAGGATATGTGGGATTTACTACAGAAATGAAA 1130
QY 960 cgttaagacttttctgctgtgttaagacacatactctcctgatgagcatttggc 1019
Db 1131 AATCCAAAAGTTGATGAGTGGGACAAAGACACATACGCCCTGATAGTATCTGCGC 1190
QY 1020 taccttgatcggtttccaggaataaccctgggaat---ttccagatccagccagatgc 1076
Db 1191 CACCATCCAAAGATTCCTGTAAGTCCGCGCTCACTCCCTCCGACCAATAAGTATGATCT 1250
QY 1077 gctgatctgaagtaagactgcctgtcaagtggaatgaatgaagcctttt--- 1133
Db 1251 ATCTGACATGCAAGACAGTTCGAGGTTGTCAAGTGGCAAGTACTTTAGGGTATGTTTC 1310
QY 1134 -----ctatcccaatttactgatatccaccttgcgaagcgtgtatttgg 1181
Db 1311 CAAGGGTGCCTTACCCGCTCGATGAGTGCATGTCGTCACATGTCGATGTCGATTTTCG 1370
QY 1182 agctgcagaattaaagtggtcttcaagaagatgacatgtgttgcataaatttgatc 1241
Db 1371 AGCTGGGACTTGAACGATGAGTGCAGCCCAACACCACTTGTGTCCTCCAAATTAAGTTC 1430
QY 1242 taagtgagccctatttgaattaaatgcttggcaga 1277
Db 1431 GGATGTTGACCTCTTGGCCATCCAGTGTGATGATGA 1466

```

```

RESULT 8
US-08-472-482-3
; Sequence 3 Application US/08472482
; Patent No. 5658778
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETA1-6
; TITLE OF INVENTION: N-ACETYLGLUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,

```

```

; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMATIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,482
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1913..1918
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION: /standard name=
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-472-482-3

```

```

Query Match 12.6%; Score 172; DB 1; Length 2105;
Best Local Similarity 51.7%; Pred. No. 4,1e-39;
Matches 515; Conservative 0; Mismatches 445; Indels 36; Gaps 4;

```

```

QY 300 tctatgtgtgtggtggaatgacagtgatgtgacatttccagactctaaagggtatgc 359
Db 489 TGACGCTATATTAACATGACAGTGCCTGTTCTTTTCATCAAGAGACGCAAAATATAT 548
QY 360 tcaaaagctgtctcaaaaggaggaagagcttcccaatagctcttctgtgttcca 419
Db 549 TGTAGAAGCCCTTAGTAAGAAAGAGGCGGAGTTTCCAAATAGCATATTCATAGTGTCTCA 608
QY 420 caaagatgaatattatggtgaaaggcttatccatgctatatcaaacagccatattta 479
Db 609 TCACAAGATGTAATGCTTGACAGGCTCTGAGGCCATATATAGCCTCAGATTTTCTA 668
QY 480 ctgcatcattatgctgtaaggacccgtataaccttcaagttgccatgaaataggc 539
Db 669 TTGCGTTATGTTGACACAAAATCCGAGGATTTCTATTACTGACGATGAGTGGCCATCGC 728
QY 540 taagtgcttcccaattttccatgcttcccaattgaagcgtgtggaatagcccaat 599
Db 729 TTCCGTTTATGATATGCTTTTGCCAGCCGATGAGAGTGGTGTATGATGCTG 788

```

```

OY 600 ttccagactccagcgtatgtaaatgtctgtcgcacctctcgaagcttcaatccagtg 659
DB 789 GAGCGGGGTGAGGCTGACCTCAACTGATGAGGATCTCTATGCAATGAGTGAACCTG 848
OY 660 gaataatgttcaactgtgtgtgcaagaatttccctgaagcgaatttgaattgt 719
DB 849 GAGAGCTTGTAAATCTTTTGGATGATGATTTCCCTTAAACCAACCTGAATGAT 908
OY 720 gtcaagatgtaaaaaactcaatgagcaaatatgttgaagacggtgaaccccaacag 779
DB 909 CAGGAAGCTCAAGTGTATAAGGAGAAACCACTGGAACGAGAGATGATGATGATGAT 968
OY 780 taatgtgaagaatcattcaccatcgaactcgaagcgtgtgacctatgtatgtgaa 839
DB 969 TAAAGAAAGAGGTGAAGAGCGGATGA-----GGTGTATTAAGAAAGCGTGC 1019
OY 840 gctacaaataaggaacaacatctccaagaagacccccccataacattcagatattgt 899
DB 1020 -----AAACACAGGAGCTGTCAAAATGCTTCTCTCACTCGAACACCTCTTTTC 1070
OY 900 tggcagctctatttctgttttaagcgaactgtgttaatatatttcaaacctccat 959
DB 1071 TGGCAGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1130
OY 960 cgtcaagacttttctgtctgtcgaagacacatactctcgtatgagcacttttgag 1019
DB 1131 AATCCAAAAGTTGATGAGTGTGCGCACAGACACATACAGCCCTGATGATGATGATGAT 1190
OY 1020 taactgtatcgtgtcgaagaatcctgtggagat---ttccagatcagccagatgt 1076
DB 1191 CACCAATCCAAAAGATCTCTGAAGTCCCGGCTCACTCCCTCCAGCCATTAAGTATGATCT 1250
OY 1077 gctcatcgtcgaagtaagactcgcctgtcgaagtggaattaccatgaagcctttt--- 1133
DB 1251 ATCTGACATGCAAGACAGTGTGCGACAGTGTGCTCAAGTGTGCTGCTGCTGCTGCTGCT 1310
OY 1134 -----ctaccagatgtcactgtatcctcgaagcgtgtgtattatgtg 1181
DB 1311 CAAAGGTGTCTCCCTACCGCCCTGCGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 1370
OY 1182 agctcgaagaattgaagtgctatcaagaatgagactcgtgtgtgttaataattgtatc 1241
DB 1371 AGCTGTGATGCTGAGTGTGATGCTGCGAACAACCACTGTTTGGCAATAAGTTTGAAGT 1430
OY 1242 taagtgagaccatcctgtatgaattgaattgtgacaga 1277
DB 1431 GGATGTGACCTCTTGGCATCCAGTGTGATGA 1466

RESULT 9
US-08-487-069-3
; Sequence 3, Application US/08487069
; Patent No. 3684134
; GENERAL INFORMATION:
; APPLICANT: FUKUDA, MINORU
; APPLICANT: BIERHUIZEN, MARTI FA
; TITLE OF INVENTION: A NOVEL BETA1-6
; TITLE OF INVENTION: N-ACETYLDIUCOSAMINYLTRANSFERASE, ITS ACCEPTOR MOLECULE,
; TITLE OF INVENTION: LEUKOSIALIN AND A METHOD FOR CLONING PROTEINS HAVING
; TITLE OF INVENTION: ENZYMACTIC ACTIVITY
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESS: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,069
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/955,041
; FILING DATE: 01-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CAMPBELL, CATHERYN
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9294
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-9001
; TELEFAX: 619-535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 220..1504
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1913..1918
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 248..314
; OTHER INFORMATION:
; OTHER INFORMATION: "SIGNAL/MEMBRANE-ANCHORING DOMAIN"
US-08-487-069-3

Query Match 12.6%; Score 172; DB 1; Length 2105;
Best Local Similarity 51.7%; Pred. No. 4,1e-39;
Matches 515; Conservative 0; Mismatches 445; Indels 36; Gaps 4;

OY 300 tgaatggtgtgtgcaatgacagcagtgatgtgacattatcagaactcaagaagtgatgc 359
DB 489 TGAAGCATATATTAACATGACAGTGTCTTCTTCTTCAATGAAGACGCAATATAT 548
OY 360 tcaaaagctgtctcacaagagagaaagcttcccaatagacctattctgtgtgtc 419
DB 549 TGTAGAACCCTTGTATTAAGAAAGAGAGGCGGAGTTTCCAAATAGCATATTTAGTGTG 608
OY 420 caaagatgaattatgtgtgaagagcttaccatgctatataacacagcaaatatta 479
DB 609 TCACAAAGATTGAATGCTTTCACAGGCTGCTGAGGCGCATCTATATGCTTCAGATTTCTA 668
OY 480 ctgacatcattatgactgaagcaccctgatacctcacaagtgtcgaatgaacaatttagc 539
DB 669 TTGCGTTCATGTGACACAAATCCGAGATTCCTATTAGCTGAGTATGGGCAATCCG 728
OY 540 taagtgcttccaatatttcaatttcttccatcaaatgaagagcgtgtgaatgacacat 599
DB 729 TTCTGTTTATGATATGCTTGTGTGCGACCGCATTTGGAGAGTGTGTTTATGATGCTG 788
OY 600 ttccagactccagcgtatgtaaatgtctgtcgcacctctcgaagcttcaatccagtg 659
DB 789 GAGCGGGGTGAGGCTGACCTCAACTGATGAGGATCTCTATGCAATGAGTGAACCTG 848
OY 660 gaataatgttcaactgtgtgtgcaagaatttccctgaagcgaatttgaattgt 719
DB 849 GAGAGCTTGTAAATCTTTTGGATGATGATTTCCCTTAAACCAACCTGAATGAT 908
OY 720 gtcaagatgtaaaaaactcaatgagcaaatatgttgaagacggtgaaccccaacag 779
DB 909 CAGGAAGCTCAAGTGTATAAGGAGAAACCACTGGAACGAGAGATGATGATGATGAT 968
OY 780 taatgtgaagaatcattcaccatcgaactcgaagcgtgtgacctatgtatgtgaa 839

```



```

Db 969 TAAAGAGAAAGGTGGAAGAAGCGGTATGA-----GGTCGTTAATGGAAGACGAC 1019
QY 840 gctaacaaataagacaacatcctccaagaagacccccccataacatcagatatgt 899
Db 1020 -----AAACACAGAGAGACTGTCAAAAATGCTTCCTCCATCGAAGACCTCTTTTC 1070
QY 900 tggcagtgctatttctgttttaagtcacgcatctgttaataatatcttcaaacatccat 959
Db 1071 TGGCAGTCTCCTACTTCGTGTCAGTAGGAGATATGTGGGTGATGACTACAGATGAATA 1130
QY 960 cgtcaagaacttttgcctgtctcaagaacacatctctctctatgagccctttggc 1019
Db 1131 AATCCAAAAGTTGATGAGATGGGCGACAGACATACAGCCCTATGATCTCTGAGGC 1190
QY 1020 taccttgatcgggtccagaataacctgagagat---ttccagatcagccagatgt 1076
Db 1191 CACCATCCAAAGGATTCCTGGAAGTCCCGGCTCCTCCTCCAGCATATGATGCT 1250
QY 1077 gtctgactcagagtaagactcgcctgtcaagtgtaagtaactatgaagctttt--- 1133
Db 1251 ATCTGACATGCAAGCAGTGTCCAGATTGTCAAGTGGAGTACTTGAAGGTGATGTTTC 1310
QY 1134 -----ctatcccaagttgtactgatactcactcgaagcgtgtgtattatg 1181
Db 1311 CAAGGTCCTCCCTACCCGCTCGATGAGATGCATGTGCGTCAGTGCATTTTCG 1370
QY 1182 agctgcgaatttaagtggtctatcaagaatgagcatgtgttctataaattgtatc 1241
Db 1371 AGCTGGTACTTGAACTGGATGTGCGCAACACCACTGTTGCCAATAAGTTTGACGT 1430
QY 1242 taagtgagccctactctgtataatgtctgacaga 1277
Db 1431 GGATGTTGACCTCTTGCCATCCAGTGTGTGATGA 1466

```

RESULT 10

US-09-063-237-3
Sequence 3, Application US/09063237
Patent No. 6124267

GENERAL INFORMATION:

APPLICANT: Mcever, Rodger P.
APPLICANT: Cummings, Richard D.
TITLE OF INVENTION: O-glycan Inhibitors of Selectin Mediated
TITLE OF INVENTION: Inflammation Derived from PSL-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30306-3450

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,237
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/649,802
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF110CIP7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-063-237-3

```

Query Match 11.8%; Score 160.8; DB 3; Length 2102;
Best Local Similarity 51.0%; Pred. No. 5,9e-36;
Matches 508; Conservative 0; Mismatches 452; Indels 36; Gaps 4;

```

QY 300 tgaatgltgttgcaatgacacagatgattgtgacattatcaagacttaagaggtatgc 359
Db 488 TGACGACTATATTAACATGACAGTGCAGTGTCTTTCATCAAGACGCAATATAT 547
QY 360 tcaaaagcttgctcaaaagagaaagcttcccaatagcctatcttgggtgcc 419
Db 548 TGTAGAACCCTTAGTAAAGAAAGAGCGGAGTTTCCAAATAGCATATTTATGTGTTCA 607
QY 420 caaagatgcaatattgttgaaggtttatccatgatatataaacagacaaattta 479
Db 608 TCAAGATGTAATGCTTGCACAGGCTGCTGAGAGGCATCTATATCCTCAGATTTCTA 667
QY 480 ctgacatcattatgactgaagacactgataccttcaaaagttgccaatgaatagg 539
Db 668 TTCCGTTCAATGTGACACAAAATCCAGATTTCTTATGATGCGCATGATGGCATCG 727
QY 540 taagtcttcccaatttcatcttcccaatgaagagctgtgaaatagcccaact 599
Db 728 TTCTGTTTATGATGCTCTTGTGCGCAGCCGATTTGAGATGTGTTATGATCGTG 787
QY 600 ttccagaactcaggtctgaatttaattgtctgtcgaacctctgaagtttcaacagtg 659
Db 788 GAGCCGGTTCAAGCTGACCTCACTGATGAAGATCTCTATGCAATGAGTCAAACTG 847
QY 660 gaatatgtatacaatttgtgtgcaagatttcccccgaagtcgaatttgaatttgt 719
Db 848 GAGTACTTGATTAATCTTTTGTGATGATTTTCCATTAAACCACTTGAAATTT 907
QY 720 gtaagaattgaaataaactaatgagcaaatatgttgaagcgttgaaccccccaag 779
Db 908 CACCAAGCTCAAGTTGTATGAGGACAAAACAACCTGGAAGAGATGCATCCCA 967
QY 780 taattggaagattcacttaccatcatgaacttagaagcgtgtccttagaatagtgaa 839
Db 968 TAAAGAACAAAGGTGAAGAGCCCTATGA-----GTCGTTAATCGAAAGCTG-- 1016
QY 840 gctaccaataagacaacatctccaagaagacccccccataacatcagatatgt 899
Db 1017 -----ACAAACACAGGAGCTGTCAAAATGCTTCTCCACACGAAACCTCTTTTC 1069
QY 900 tggcagtgctatttctgttttaagtcacgcatctgttaataatatcttcaaacatccat 959
Db 1070 TGGCAGTCTCCTACTTCGTGCTCAGTAGGAGATATGTGGGTGATGACTACAGATGAATA 1129
QY 960 cgtcaagaacttttgcctgtctcaagaacacatctctctctatgagccctttggc 1019
Db 1130 AATCCAAAAGTTGATGAGATGGGCGACAGACATACAGCCCTATGATCTCTGAGGC 1189
QY 1020 taccttgatcgggtccagaataacctgagagat---ttccagatcagccagatgt 1076
Db 1190 CACCATCCAAAGGATTCCTGGAAGTCCCGGCTCCTCCTCCAGCATATGATGCT 1249
QY 1077 gtctgactcagagtaagactcgcctgtcaagtgtaagtaactatgaagctttt--- 1133
Db 1250 ATCTGACATGCAAGCAGTGTCCAGATTGTCAAGTGGAGTACTTGAAGGTGATGTTTC 1309
QY 1134 -----ctatcccaagttgtactgataccttgaagcgtgtgtattatg 1181
Db 1310 CAAGGTCCTCCCTACCCGCTCGATGAGATGCATGTGCGCTCAGTGCATTTTCG 1369

```


Db 13 CCTTATCTAAGAAGAAAGCTGACTTCCCTTGGCATATATATAGTCATCATCATCAC 72
QY 427 gcaattatggttgaaggcttatccatgctatatacaaccgagcaattatgctatc 486
Db 73 TTGACACCTTTGCAAGGCTCTTCAAGGCTATTACATGCCCAAAATATCTACTGTGT 132
QY 487 cattatgctgtaaggcaccctgataccttcaaaagtgtccatgaaacaaattagctaaagtc 546
Db 133 CATGTGATGAAAAAGCAACAAGTAATTTAAAGATGGGGTAGAGCAACTATTAAAGCTGC 192
QY 547 ttcccaatatttcatgctgtcccaaatltagaggtgtgtgaatatgtcccaattccaga 606
Db 193 TTCCCAAAACGCTTTTGTGCTTCCCAAGATGGAACCCGTGTCTATGAGAGGATCTCCAGG 252
QY 607 ctccaggtgatttaaatgtctgtcgcacctctcgaagtcctcaatccagtggaatat 666
Db 253 CTCACAGGCTGACCTGAAGTGCATCAGAGATCTTCTGCTTCCAGGTCATGAGAACTAC 312
QY 667 gtatacaactgtgtggcaagatttccctcgaagtcgaatttgaattgtgtcagag 726
Db 313 GTATCAACACCTGTGGGCAAGACTTCCCTGAAAAACAAGAAATAGTTCAGTAT 372
QY 727 ttgaaa 732
Db 373 CTGAAA 378

RESULT 13
US-08-488-135-1
Sequence 1, Application US/08488135
Patent No. 5766910
GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,135
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS

LOCATION: 1..378
US-08-488-135-1
Query Match 8.9%; Score 121.2; DB 1; Length 378;
Best Local Similarity 58.2%; Pred. No. 4.2e-25;
Matches 213; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 367 ctgtgtccaaggaggaagaagcttcccaatagctatcttctgtgtccacaagaat 426
Db 13 CCTTATCTAAGAAGAAAGCTGACTTCCCTTGGCATATATATAGTCATCATCATCAC 72
QY 427 gcaattatggttgaaggcttatccatgctatatacaaccgagcaattatgctatc 486
Db 73 TTGACACCTTTGCAAGGCTCTTCAAGGCTATTACATGCCCAAAATATCTACTGTGT 132
QY 487 cattatgctgtaaggcaccctgataccttcaaaagtgtccatgaaacaaattagctaaagtc 546
Db 133 CATGTGATGAAAAAGCAACAAGTAATTTAAAGATGGGGTAGAGCAACTATTAAAGCTGC 192
QY 547 ttcccaatatttcatgctgtcccaaatltagaggtgtgtgaatatgtcccaattccaga 606
Db 193 TTCCCAAAACGCTTTTGTGCTTCCCAAGATGGAACCCGTGTCTATGAGAGGATCTCCAGG 252
QY 607 ctccaggtgatttaaatgtctgtcgcacctctcgaagtcctcaatccagtggaatat 666
Db 253 CTCACAGGCTGACCTGAAGTGCATCAGAGATCTTCTGCTTCCAGGTCATGAGAACTAC 312
QY 667 gtatacaactgtgtggcaagatttccctcgaagtcgaatttgaattgtgtcagag 726
Db 313 GTATCAACACCTGTGGGCAAGACTTCCCTGAAAAACAAGAAATAGTTCAGTAT 372
QY 727 ttgaaa 732
Db 373 CTGAAA 378

RESULT 14
US-08-474-065-1
Sequence 1, Application US/08474065
Patent No. 5830465
GENERAL INFORMATION:
APPLICANT: Fukuda, Minoru
TITLE OF INVENTION: Expression of the Developmental I
TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
NUMBER OF SEQUENCES: 14
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,065
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/118,906
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 9526
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 378 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..378
 US-08-474-065-1

Query Match 8.9%; Score 121.2; DB 2; Length 378;
 Best Local Similarity 58.2%; Pred. No. 4.2e-25;
 Matches 213; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Db 367 ctgtctcaagaagaggaagaaagcttcccaatagcctattcttgggtggtcccaagaat 426
 13 CCTTATCTAAGAGAAGAGCTGACTTCCCTGGCATATATTAATGGTCATCATCATCAC 72
 QY 427 gcaatttggttgaagagctatccatgcatatatacaacagcaaatatttcgcatc 486
 73 TTTGACACCTTTGCCAAGGCTCTTCAAGGCTATTTCATGCCCCCAAAATATCTACTGT 132
 Db 487 catatgatcgtgaagacacctgataccttcaaaagtgcataagaaattagctaaagtgc 546
 133 CATGTGATGTAAGAAAGCAACAACTGAATTAAAGATGCGTAGAGCAACTATTAAAGCTGC 192
 QY 547 ttcccaaatatttcattgcttcccaattgagagctgtgtaatatggcccaatttcaga 606
 193 TTCCCAACGCTTTCTGCTGCTCCCAAGATGAGACCCGTTGCTATGAGAGGATCTCCAGG 252
 Db 607 ctccagagctgattaaattgctgtcgcacctctgaagcttcaatccatccagttgaaatat 666
 253 CTCAGGCTGACGTAAGTGCATGAGATCTTCTGCTCGAGGCTCATGGAAGTAC 312
 QY 667 atttcaactcgtgtggaagaaatttccctgaagaaatatttgatgtgtcagag 726
 313 GTATCAACACCTGTGGCAAGACTTCCCTGAAACCAACAGGAATATGTTCACTAT 372
 Db 727 ttgaag 732
 373 CTGAAG 378

RESULT 15
 US-08-118-906-3

Sequence 3, Application US/08118906
 Patent No. 5484590

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

APPLICANT: Blerhuizen, Marti F.A.

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a

TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,906

FILING DATE: 09-SEP-1993

CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 9526
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 378 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..378
 US-08-118-906-3

Query Match 6.8%; Score 92.2; DB 1; Length 378;
 Best Local Similarity 52.8%; Pred. No. 6.3e-17;
 Matches 199; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 355 tatgtccaaagcttgctccaagaaggaagaaagcttcccaatagcctattcttgggt 414
 1 TATATTGTGAACCCCTTAGTAAGAAGAGCGGAGTTCCAAATGCAATTTCTATATGTG 60
 Db 415 gtccacaagaatgcaattatgttgaagagctatccatgcatataaacaagacaat 474
 61 GTTCATCAACAAGTTGAATGCTTGACAGCTGTGAGGCGCATCTATATGCTCAGAAAT 120
 QY 475 atttactgcatcattatgataatgtaaggaacctgataccttcaagttgcatgacaat 534
 121 TTCTATTGGTTCTATGTGACACAAATCCGAGGATTCCTATTAGCTGACGTGAGGCG 180
 Db 535 tttagcgaagtgctctccaatatatttcattgcttcccaattagagagctgtgaaatatgc 594
 181 ATGCGTTCCTGTTTAGTATGTCCTTTGTGGCCAGCCATGTGAGAGTGTGTTATGCA 240
 QY 595 cacatttceagactccaagctgatttcaattgctgtggaacctctgaagcttccatc 654
 241 TCGTGAAGCCGGGTTCAGGCTGACCTCAACTGCATGAAGATCTATGCAATGAGTGCA 300
 Db 655 cagtgaataatgttatacaactgtgtggaagaaatttccctgaagtcacaatttgaa 714
 301 AACTGGAAGTACTGTATAAATCTTTGTGTATGATTTCCCATTAACCAACCACTAGAA 360
 QY 715 ttggtgtcagagttgaa 731
 361 ATTGTCAAGAAAGCTCAA 377

Search completed: May 2, 2002, 10:30:28
 Job time: 7703 sec

Fri May 3 10:57:52 2002

us-09-645-192-1.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 08:20:30 ; Search time 1980.83 seconds

(without alignments)
11343.305 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362

Sequence: 1 atgaagatatacaatgta.....attcactacacacatga 1362

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8246589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_in.*
5: gb_ov.*
6: gb_ov.*
7: gb_ph.*
8: gb_ph.*
9: gb_ph.*
10: gb_ph.*
11: gb_ph.*
12: gb_ph.*
13: gb_ph.*
14: gb_ph.*
15: gb_ph.*
16: gb_ph.*
17: gb_ph.*
18: gb_ph.*
19: gb_ph.*
20: gb_ph.*
21: gb_ph.*
22: gb_ph.*
23: gb_ph.*
24: gb_ph.*
25: gb_ph.*
26: gb_ph.*
27: gb_ph.*
28: gb_ph.*
29: gb_ph.*
30: gb_ph.*
31: gb_ph.*
32: gb_ph.*
33: gb_ph.*
34: gb_ph.*
35: gb_ph.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	6 AX087935	AX087935 Sequence
2	1362	100.0	3435	9 AF132035	AF132035 Homo sapi
3	1358.8	99.8	184590	2 AC093259	AC093259 Homo sapi
4	207.2	15.2	3508	10 AB037596	AB037596 Mus muscu
5	194.8	14.3	108873	14 AF18573	AF18573 Bovine he
6	194.4	14.3	1206	10 AB037597	AB037597 Mus muscu
7	191.8	14.1	1317	6 AX087948	AX087948 Sequence
8	191.8	14.1	1317	9 AF038650	AF038650 Homo sapi
9	191.8	14.1	2128	6 AR136078	AR136078 Sequence
10	191.8	14.1	2128	9 AF102542	AF102542 Homo sapi
11	191.8	14.1	2216	6 AX045249	AX045249 Sequence
12	191.8	14.1	181983	33 AC022480	AC022480 Homo sapi
13	191.8	14.1	187760	2 AC092755	AC092755 Homo sapi
14	191.6	14.1	2017	14 AF231105	AF231105 Bovine he
15	191.2	14.0	2393	10 MM068182	MM068182 Mus musculu
16	187.4	13.8	87636	9 AL139039	AL139039 Homo sapi
17	187.4	13.8	176544	2 AL138877	AL138877 Homo sapi
18	187.4	13.8	196788	2 AC021626	AC021626 Homo sapi
19	185.2	13.6	1203	6 AX087950	AX087950 Sequence
20	185.2	13.6	1807	6 I15889	I15889 Sequence 13
21	185.2	13.6	1807	6 I15889	I15889 Sequence 13
22	185.2	13.6	1807	6 I15889	I15889 Sequence 13
23	185.2	13.6	1807	6 I15889	I15889 Sequence 13
24	184.2	13.5	157102	2 AC076970	AC076970 Homo sapi
25	183.6	13.5	155359	9 AL358777	AL358777 Homo sapi
26	183.6	13.5	196788	2 AC021626	AC021626 Homo sapi
27	181.2	13.3	252504	2 AC073815	AC073815 Mus muscu
28	180.8	13.3	2217	6 AX045251	AX045251 Sequence
29	180.8	13.3	2217	6 AX045251	AX045251 Sequence
30	179.6	13.2	2033	10 MM019265	MM019265 Mus musculu
31	175.2	12.9	159930	2 AL138757	AL138757 Homo sapi
32	175.2	12.9	164041	2 AL590664	AL590664 Homo sapi
33	175.2	12.9	183858	9 AL161626	AL161626 Homo sapi
34	174.4	12.8	5030	10 S79797	S79797 enzymatic g
35	174.4	12.8	2060	9 H51671	H51671 Human Beta
36	173.2	12.7	2023	10 D87333	D87333 Mus musculu
37	173.2	12.7	2047	10 D87332	D87332 Mus musculu
38	172.8	12.7	4179	14 BHV4RBEF	BHV4RBEF Bovine herp
39	172	12.6	1287	6 AX087946	AX087946 Sequence
40	172	12.6	2105	6 I41262	I41262 Sequence 3
41	172	12.6	2105	6 I61440	I61440 Sequence 3
42	172	12.6	2105	6 I72548	I72548 Sequence 3
43	172	12.6	2110	6 HUMGOGG	HUMGOGG Human Beta
44	172	12.6	2204	9 HUMC2CNT	HUMC2CNT Homo sapien
45	164.8	12.1	2435	4 BTU41320	BTU41320 Bos taurus

ALIGNMENTS

RESULT 1
AX087935
LOCUS AX087935
DEFINITION Sequence 1 from Patent WO0114535.
ACCESSION AX087935
VERSION AX087935.1 GI:13396913
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1362)
AUTHORS Schwientek,T. and Clausen,H.
TITLE UDP-n-acetylglucosamine: galactose--g(b)1.3-n
-acetylglucosamine--g(a)-r/(glucnac to galnac)
-g(b)1.6-n
Patent: WO 0114535-A 1 01-MAR-2001;
Schwientek, Tilo (DK) ; Clausen, Henrik (DK)
FEATURES
Source Location/Qualifiers
1..1362

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
1.1362
/note="The gene of C2GnT3"
/codon_start=1
/protein_id="CAC34674.1"
/db_xref="GI:13396914"
/translation="MKFRKCYFKHTLOOKFLLFLMLTSLIKLVNRLPRQNDI
LVEYSSTSPFVNRNTHVDEVRVENCSEITGEPLLEIGKLELRKDIDLED
VVAATSDCLYQTLRGRYAKLVSKESSEPIASLVHBDALVRLIAITVQHNII
CIHYDRAPDIFRYANNNLAKCSNIFIAKLEAVYAHISRLQADNLSDLKSSI
QMKYVIMLCODEPLKSNFELVSELEKLVANMLFVPSKLEFRTYHHELRVY
EYVLPRTNISKSEAPPHNIQIFVGSAYFVLSQAPVKYIENINSIODPFMSDTPS
DEHFMATLIVPGIPGEISRSADVSLOSRLTVMWNYEGEFPSGSHRSYCI
YGAELFMILKDGHWPFANKRDSKVDPIILNCLAEKLEEQORDMITLPSEKLEMDRNL
TTS"
BASE COUNT      410 a      260 c      279 g      413 t
ORIGIN
Query Match      100.0%; Score 1362; DB 6; Length 1362;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagataatcaatgttattttaacataccctacagcaagaatttcatcctgtt 60
Db 1 ATGAAGATATCAATGATTATTAAACATACCTACAGCAGAAAGTTTCATCCTGTT 60

QY 61 ttaaccatgagctgctcttctgttaagcttctaagtgaagagactcttcccaa 120
Db 61 TTAACCTATGAGCTCTCTCTTGTGTAAGCTTCAATGTAAGCGACCTCTTCCGAA 120

QY 121 aaagacattacttgcttgagtaactccctaaagtaacctgcttctgtlaagaagaatc 180
Db 121 AAAGACATTACTTGTTGAGTACCTCCTAAGTACCTGCTTTGTGTAAGAACAGATAC 180

QY 181 actatgttaagatgaagatcaagatgaagtaactgttctggatcttaagaacagag 240
Db 181 ACTATGTTAAAGATGAAGTCAAGGTATAGATTAAGTGTGGGATCTTAAGAACAGAG 240

QY 241 ccttgaagaattgaaagagctcgtgaataaagaagaagagacatcatgacttggagat 300
Db 241 CCTTGGAATGTGAAGAGTCTGGAATTAAGAAAGAGACATCATGACTTGGAGAT 300

QY 301 gatgtgttggcgaatgaacagtgatgtgacatttcagactcttaagaggttaagct 360
Db 301 GATGTGTTGGCGAATGACAGTGTATGACATTATACAGACTTAAGAGGTTATGCT 360

QY 361 caaaagcttgctcaaaagagagaaagcttcccaatagacctattcttggctgtccac 420
Db 361 CAAAAGCTTGCTCAAAAGAGAGAAAGCTTCCCAATAGCCTATTCTTGGTGTCCAC 420

QY 421 aaagatgcaattatggttgaagagcttcaactctatacaacaacagacacaattac 480
Db 421 AAAGATGCAATTATGTTGTAAGGCTTATCCATGCTATATCAACACCAACAATATTAC 480

QY 481 tgcacccatlatgacgtcaagcaccatgataccctcaagtgccatgaacaattagct 540
Db 481 TGCACCATLATGATCGTAAAGCACCATATACCTTCAAAAGTGGCATTAACAATTTAGCT 540

QY 541 aagtgcttcccaatatttcatgtcttccaaatagaagagctgtggaatagcccaatt 600
Db 541 AAGTCTCTCTCCAAATATTTTCATTCCTTCCAAATAGAGGCTGTGGAATAGCCCACTT 600

QY 601 tccagactcagagctgatttaattatgctgtgcgaccttcgaagcttccaatcagtg 660
Db 601 TCCAGACTCAGAGCTGATTAAATGCTGTGCGACCTTCGAGACTCTTCAATCCAGAGG 660

QY 661 aaatatgtatcaactgtgtgagcaagaatttccctcgaagtcgaatttgaattgagtg 720
Db 661 AAATATGTTATCAACTGTGTGGGCAAGATTTTCCCTCGAAGTCAAAATTTGAAATGGGTG 720

QY 721 tcagagttgaaaaaactcaatgagcaaatatgttgaagacagtgaaacccccaacagtg 780

```

```

Db 721 TCAGAGTTAAAAAACTCATATGAGCAAAATATGTTGGAACGCTAACCCTCAACAGT 780

QY 781 aaattgaaagattcaccatcatcatgaactagaacgggtgcctatgaatagtgaag 840
Db 781 AAATTGGAAGATTCACCTACCATCATATGAAGCAAGGCGGTGCTTATGAATATGTGAAG 840

QY 841 ctccaataagaacaacatctccaaagaaagacccccccataacaattcagaatttct 900
Db 841 CTACCAATAAGGCAACAACATCTCCAAAGAAAGAACCCCCCAACAACTTCAATATTGTT 900

QY 901 ggcagtgccatttctttaaagcaagcattgttaatatatttcaacaactccatc 960
Db 901 GCGAGTGCTATTATTGTTTAAAGTCAAGATTTGTTAAATATTTTTCACAACATCCATC 960

QY 961 gtccaagacttttctcctggtctaaagaacatactctcctgtatgagcactttggact 1020
Db 961 GTTCAAGACTTTTTCCTGCTCTAAAGACATACCTCTCTGATGAGCCTTTTGGGCT 1020

QY 1021 acctgattcggttccaggaataacctgggagagattccagatacagccagatgtgct 1080
Db 1021 ACCTTGATTCGGGTTCCAGGAATACCTGGGAGATTTCCAGATCACCCAGATGTGCT 1080

QY 1081 gatcgcagagtaagatcgcgcttgcgaagtgaattactatgaagagcttttctatccc 1140
Db 1081 GATCGCAGAGTAAAGACTGCGCTTGTCAAGTGAATTAAGAAAGGCTTTTCTATCCC 1140

QY 1141 agttgactgagatctacaccttgaaagcgtgtatattaaagagcgtgcgaatlaagtg 1200
Db 1141 AGTTGACTGAGATCTACCTTCAAGCGTGTGATTTAAGAGCTGCACAAATTAAGAGTG 1200

QY 1201 ctatacaagatggaacattggttcttaataaattgactcaagtggaacctatctgt 1260
Db 1201 CTATCAAGATGGAACATTGTTGCTTAATTAATTTGATTTGAAGTGAACCTTATCTTG 1260

QY 1261 attaaatgcttgcgaagaaagcttgaagaaagcagaagactgagatcacttggccctca 1320
Db 1261 ATTAATAGCTTGCAGAAAGCTTGAAGAACGACAGAGACTGATCACTTGCCCTCA 1320

QY 1321 gaaagattatlatgataagaatccacbacacacatctga 1362
Db 1321 GAAAGTATTATTATGATAGTAATCTCATTACACATCATGA 1362

RESULT 2
AF132035 3435 bp mRNA PRI 08-APR-2000
LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3
DEFINITION (C2GnT3) mRNA, complete cds.
ACCESSION AF132035
VERSION AF132035.1 GI:7527463
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3435)
AUTHORS Schwientek, T., Yeh, J. C., Levery, S. B., Keck, B., Merx, G., van
Kessel, A. G., Fukuda, M. and Clausen, H.
CONTROL OF O-glycan branch formation. Molecular cloning and
characterization of a novel thymus-associated core 2 beta1,
6-n-acetylglucosaminyltransferase
J. Biol. Chem. 275 (15), 11106-11113 (2000)
20219156
MEDLINE 2 (bases 1 to 3435)
REFERENCE Schwientek, T. and Clausen, H.
AUTHORS Direct Submission
TITLE Submitted (27-FEB-1999) School of Dentistry, University of
Copenhagen, Nørre Alle 20, Copenhagen 2200, Denmark
JOURNAL Location/Qualifiers
FEATURES
1.3435
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

gene
1. 3435
/gene="C2Gnt3"
CDS
862..2223
/gene="C2Gnt3"
/codon_start=1
/product="core 2 beta-1,6-N-acetylglucosaminyltransferase
3"
/protein_id="AA63156.1"
/db_xref="GI:7527464"
/translation="MKLFKCYFKHTLOQKVFILFLTMLSLKLNVRRLFPQKDIY
LVEYLSLSTPFRNRYTHVDEVRVENCISYOEPLKSLERRDIDLEDID
VYAMTSDCIQIOLRGLYQOKLVSKESFPLAYELVYKDAIWERLHAIVNCHNY
CIHYDRKAPDTEFKVANNKLACFSNIFLASKLEVEVAHLSRLOADNCLSDLSKI
OMKXVILNGODPFLKSNFELSVSLKLNANMLETVPKPSNLEFRFYHSELRYVY
EYKPLIRPTNSKAPPHNTOIFYGSAYFVLSQAFVYIINNLSIVQDFPMASKDTPS
DEHWATLIRPFGIRPEGLISRSADVPDQSKTRVAKNNITEGEFTPCTSHLSRSCI
YGAELFMWLKIDGHWPFANKFDSKVDPIILIKCLAEKLEEQRDWITLPESEKLFMDRNL
TTS"
BASE COUNT 1032 a 678 c 651 g 1074 t
ORIGIN

```

```

Query Match 100.0%; Score 1362; DB 9; Length 3435;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatactcaatgttattttaacataccctacagcagaaggtttcatcctgtt 60
DB 862 ATGAAGATATCAAAATGTTATTTTAAACATACCCACAGCAAGATTTCATCCTGTT 921
QY 61 ttaacctatggcgtcctcttctttaaagcttcaaatgtgagcagcctttccgaa 120
DB 922 TTAAACCCATGGCTCCTCTCTTTGTTAAAGCTTCTAAATGTGAGCGACTCTTCCGCA 981
QY 121 aaagacatttacttgctgagactccttaagtaactcgccttcttgtaagaaacagatc 180
DB 982 AAAGACATTTACTTGTTGAGACTCCCTAAGTACCTTGCTTTGTAAGAAACAGATAC 1041
QY 181 actcatgttaagatgaagtcagatgaagttaactgttcgggtatcatalgaacaggag 240
DB 1042 ACTCATGTTAAGAGTAAGTACGATGATGAAGTTACTGTGGGATATCTATGAACAGAG 1101
QY 241 ccttggaatgtgaaagagcttgaaataagaaagagacatcatgtacttgagagat 300
DB 1102 CCTTGGAAATGGAAGAGTGTGAAATTAAGAAAGAGGACATCATTTGAGGAGAT 1161
QY 301 gatgatgttgygcaatgacagatgattgtgacattatcagactctaaagaggtatgct 360
DB 1162 GATGATGTTGTGCAATGACAGTGAATTTGATGACACTCTAAGAGGTTATGCT 1221
QY 361 caaagcttgctcaaaagagagaaagcttcccaatagcctattcttgggtgcac 420
DB 1222 CAAAAGCTGTCTCAAAAGAGGAGAAAGCTCCCAATAGCCTATTCTTGGTGTCCAC 1281
QY 421 aaagatgcaattatgctgaaagcttccatgctatataccaacagcacatattac 480
DB 1282 AAAAGTGAATATGATGTTGAAGGCTTATTCATGCTATATACAAACACACATTTTAC 1341
QY 481 tgcacatcatatgatcgttaagcactgtatccttcaagtgcacatgaacaatttagct 540
DB 1342 TGCATCCATTAATGATGTAAGGACCTGATACCTTCAAAAGTGGCCATTAACAATTTACT 1401
QY 541 aagtgctctccaatatttcttgctcccaatagaggcgtgtgaaatgacccacatt 600
DB 1402 AAGTGTCTTCCAAATTTTTCATTTCTTCCAAATTAAGGCGTGTGAAATAGCCACATT 1461
QY 601 tccagactcagagctgatttaaatgtctgtgcagccttcgaagcttccaatccagtyg 660
DB 1462 TCCAGACTCCAGCTGATTAATGCTGTGCGACCTTCGAAGTCTTCAATCCAGTGG 1521
QY 661 aaatagtatacaacttgtgtgcaagaatttccctggaagtcaaatlttgaaattgtg 720
DB 1522 AAATATGTTATCACTGTGTGCGCAAGATTTCCCTGGAAGCTCAAAATTTGCAATGATG 1581

```

```

QY 721 tcaagattgaaaaaactcaatlgagcacaatctgttggagacggtgaaaccccccaacagt 780
DB 1582 TCAGAGTTGAAAAAATCAATGAGAGCAAAATATGTTGAGAGCGGTGAACCCCCCAACAGT 1641
QY 781 aaatgaaagatcacttaccatcatgaacttagacgggtgtccttatgatatgtgag 840
DB 1642 AAATGGAAGATTCACTTACCATCATGAACTTAGAGAGGGGTGCTTATGATATGTGAAG 1701
QY 841 ctaccaaagagacaacatctccaaggaagcaccocccataacatctcagatatgtct 900
DB 1702 CTACCAATTAAGAGCAAAACATCTCCAAAGAGCAACCCCCCATTAACATTCAGATATTGTT 1761
QY 901 ggcagtgctatcttgttcttaagtcaagccttggtaaatatatttcaaacatccatc 960
DB 1762 GCGAGTCTTATTTGTTTAAAGTCAAGCAAGCTTTGTTAAATATATTTTCAACAACCTCATT 1821
QY 961 gtccaagacttlttgcttgctcgtcctaaagacatactctcctgatgacacttggact 1020
DB 1822 GTTCAAGACTTTTTCCTGCTGCTGTAAGACACATACCTCTGATGAGCACTTTTGGGCT 1881
QY 1021 acctgattcgggttccagaataacctggtgagatttccagatcagcccaagatgtgct 1080
DB 1882 ACCTTGATTTGGGTTCCAGGAATACCTGGGAGATTTCCAGATCAACCCAGATGTGCT 1941
QY 1081 gatctcagagtaagactcgccttgcaggttggaattactatgaaggtttttctatcc 1140
DB 1942 GATCTGAGAGTAAAGACTCCCTTGTCATAGTGAATTAAGGCTTTTCTATCC 2001
QY 1141 agttgactgatactcaacttccgaagcggtgtatttaagagctcgaagaattaagtyg 1200
DB 2002 AGTTGACTGATCTCACTCACTTCGAAGCGTGTGATTTATGAGCTCAGATTAAGTGG 2061
QY 1201 cttaacaaagatgagacttggttgcataaattgattcgaagttgagcccatcttg 1260
DB 2062 CTTTCAAAAGTGAACATTTGCTTTCCTAATAATTTGATTTAAGTGAACCTTATCTTG 2121
QY 1261 attaatgcttggagaaagacttgaagaaacagcagagagactgtatcaacttgcctca 1320
DB 2122 ATTAAATGCTTGGCAGAAAGCTTGAAGAACACAGAGAGACTGATCACTTGGCCCTCA 2181
QY 1321 gaaagcttattatgataagaatctcactaacacatcaga 1362
DB 2182 GAAAGTATATTATGATGAAGAAATCTCACTACCAACATCATGA 2223

RESULT 3
AC093259/c
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone RP11-229c3, WORKING DRAFT SEQUENCE,
6 unordered pieces.
AC093259
VERSION
AC093259.1 GI:15193393
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 184590)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
PUBLISHED
2 (bases 1 to 184590)
REFERENCE
DOE Joint Genome Institute.
AUTHORS
Direct Submission
TITLE
DOE Joint Genome Institute.
JOURNAL
Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information

```

Center Project Name: 494352
Center clone name: RPCI-1L_229C3

Summary Statistics

Consensus quality: 176980 bases at least Q40
Consensus quality: 182494 bases at least Q30
Consensus quality: 182978 bases at least Q20
Estimated insert size: 183600; agarose-ef estimation
Estimated insert size: 184090; sum-of-ctnigs estimation
Quality coverage: 8.9 in Q20 bases; agarose-ef estimation
Quality coverage: 8.9 in Q20 bases; sum-of-ctnigs estimation.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2136: contig of 2136 bp in length
* 2137 2236: gap of unknown length
* 2237 5820: contig of 3584 bp in length
* 5821 5920: gap of unknown length
* 5921 8425: contig of 2505 bp in length
* 8426 8525: gap of unknown length
* 42924: contig of 3439 bp in length
* 42925 43024: gap of unknown length
* 43025 98717: contig of 55693 bp in length
* 98718 98817: gap of unknown length
* 98818 184590: contig of 85773 bp in length.

FEATURES

SOURCE

1.184590
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-229C3"
/clone_id="RPCI human BAC library 11"

BASE COUNT 56307 a 36971 c 36814 g 53956 t 502 others
ORIGIN

Query Match 99.8%; Score 1358.8; DB 2; Length 184590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatatacaatgttatatttaaacatccctacagcagaagtttcatccgtttc 60
Db 22472 ATGAAGATATTCAAATGTATTTTAAACATACCCCTACGACGAAAGTTTCATCCTGTTT 22413
QY 61 ttaacctatggtcgtctctcttctttaaagcttctaagtgtgagcagactcttcgcaa 120
Db 22412 TTAACCCATGCGTCTCTCTTTGTTAAAGCTTCTAAATGTGAACGACGCTTTCCGCA 22353
QY 121 aagagcattactggtgtgaagtaaccttaagtaacctgctcttcttgaagaacagatac 180
Db 22412 TTAACCCATGCGTCTCTCTTTGTTAAAGCTTCTAAATGTGAACGACGCTTTCCGCA 22353
QY 22332 AAAGCATTCTTCTGTTGTAAGTACTCCCTAAGTACTCGCTTTGTAAGAAAGATAC 22293
QY 181 actcatgttaagatgaagtcaggtatgaagtaagtaagtaagtaagtaagtaagtaag 240
Db 22292 ACTCATGTTAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 22233
QY 241 cctttggaatgtgaagaagtcctggaataaagaagaagaagaagaagaagaagaaga 300
Db 22232 CCTTTGGAATGTGGAAGAAGTCTGGAATAAGAAAGGACATCTGACTTGAGAGAT 22173
QY 301 gatgatgtgtgcaatgacagtgatgtgacattatcagactcctcaagaagtgatgct 360
Db 22172 GATGATGTGTGCGCATACAGTGAATGTGATGATGATGATGATGATGATGATGATG 22113
QY 361 caaaagctgtctcaagaagagagaaagcttcccaatagcctatctcttggtgtgtccac 420
Db 22112 CAAAAGCTTGTCTCAAGAGAGAGAAAGCTTCCCAATAGCCTATCTTGTGTTGTCCAC 22053

QY 421 aagaatgcaattatgtttgaaaggtctatccatgctatatataaccagacataattac 480
Db 22052 AAAGATGCAATATATGTTGTAAGGCTTATCCATGCTATATACACACGACATATTATAC 21993
QY 481 tgcattccattatgatcgttaagagacctgtataccttaaaagttgcataagcaattagct 540
Db 21992 TGCATTCATTATGATGCTATGAGGACCGCTGATACCTTAAAGTTGCCATGAAACATTAAGCT 21933
QY 541 aagtgcttcccaataatttccattgcttcccaattgagagcgtgtggaatatgcccacat 600
Db 21932 AAGTGCTTCCCAATATTTTCATTGCTTCCAAATTAAGAGCGTGTGAAATATGCCACATT 21873
QY 601 tccagactccagagctatatttaattgtctgtgagacctctgagactcctcaatccagtg 660
Db 21872 TCCAGACTCCAGGCTGATTTAAATGCTGTGTGAGACCTTGTGAAGCTTCATCCAGTGTG 21813
QY 661 aaataatgttcaactctgtgtggaagaatttccctgaagtcataaatttgaattgtg 720
Db 21812 AAATATGTTATACAACTTGTGTGTGGCAAGATTTTCCCTGAAGTCAAAATTTGAATTGGTG 21753
QY 721 tcaagatgtgaaaaaactccaatgtgagcaaatatgttgagagcgttgagaccccaaacag 780
Db 21752 TCAGAGTTGAAAAAACTCAATGAGCAATATATTTGGAAGCGTGAACCCCAACAGCT 21693
QY 781 aaattggaagatccacttaaccatcaatgaactagaacggtgtgctatgaatatgtgaag 840
Db 21692 AAATTGGAAGATTCCTACTTACCATCATGAACCTAGACGGGTGCTTATGAATATGGAAG 21633
QY 841 ctcccaataagagcaaacacttcccaagaagaagaccccccaataacatgaatttgtt 900
Db 21632 CTACCATTAAGGACAAACATCTCCAAAGAGACCCCCCAATPAACCTTGAATATTTGTT 21573
QY 901 ggcagtgctatttctttaaagcaacatctgttaataatatttcaaacactccalc 960
Db 21572 GGCAGTGCTATTTTGTGTTTAAAGTCAAGCATTTGTTAAATATTTTCAACAACTCCATC 21513
QY 961 gtccaagactcttttccctggtcttaagaacacactctccctgaagagacatttgggct 1020
Db 21512 GTTCAAGACTTTTTCCTGGTCTTAAGACACATCTCCGATGACGACCTTTGGGCT 21453
QY 1021 acctgattcgggttccagaagaatccctgggagatttccagatccagccagatgtgct 1080
Db 21452 ACCTGATTTCGGTTCCAGGAATACCTGGGAGATTTCCAGATCCAGCCAGATGTGCT 21393
QY 1081 gactcagaagtaagaactgctgttcaagtgaattactatgaagcttttctatccc 1140
Db 21392 GATCTCAGAGTGAAGTCTGCTTGTCAAGTGAATTAATGAAGGCTTTTCTATCCC 21333
QY 1141 agttgacagatcagatccactcgaagcgtgtgatttcaagagctcgaagtaaggtg 1200
Db 21332 AGTTGTAAGTGAAGTCACTGCAAGCGTGTGATTTATGAGAGCTGCACAAATTAAGTGTG 21273
QY 1201 ctatccaaagatgagacattggttttctaaataattgattcctaaggtgagacctatctg 1260
Db 21272 CTTATCAAGAGTGAAGTATGTTGCTAATTAATTTGATTTCAAGTGAAGCCTTATCTTG 21213
QY 1261 attaaatgcttgcagaagaagcttgaagaagaagcagaagaagatgataccttgcctca 1320
Db 21212 ATTAATGCTGTGCAAAAAAGCTGAAGAACAGACAGAGATGATGATCTTGGCCTCA 21153
QY 1321 gaaaagtattatgtagtagaatactactaccacatcga 1362
Db 21152 GAAAAGTTATTATGATGATGAATACTACCTACCATCATGA 21111

RESULT 4
AB037596 3508 bp mRNA ROD 20-Oct-2000
LOCUS AB037596
DEFINITION Mus musculus Ignt B mRNA for
ACCESSION AB037596
VERSION AB037596.1 GI:9650953
KEYWORDS beta-1,6-N-acetylglucosaminyltransferase B.

PIPGFYNTPLVYDFASLYPSIIQAHNLQYSTMIDONLHLHLKPDYETHELSTGP
 IHFKOHKTKSLSTLITLTAAMAKRTIRELANCDGPKRTILDKOOLIKYTCNSVY
 GFTGVAISLIPICPIAEVYVLOGRTMLEKSEKAVENIPEERISDIYSYVPPDPASV
 RVIYGDITSLIECRGYPMSOVLNFPADKLETTETTKLEFKLEKFOCLIMTK
 KRYIGLSSKXIKVMKGVDLIRKACSFNOTITSEKIDILRDPREKQAQYICRQAP
 KYSDSLPVGFKYIDVYLNNSYALATGKVSSELSFELSSEFVEYITTLPLHLY
 YRKINSREELPDHRIPIYVFKDQCKSDLDPLVYSONKIPISVETIFDKLI
 HGVALIQLLEFGNSMTEILYNFVINYEST"
 17297. 18577
 /note="ORF 10: B0FB1; conserved in other
 gamma-herpesviruses"
 /codon_start=1
 /product="unknown"
 /protein_id="AK07929.1"
 /db_xref="GI:12802537"
 /translation="MTSLVEFSIEGMQVNICNGIILMTNLKIDIPVQGYMLVKV
 FLPSLNOLINNYISFLVOSIEFEDYACQAMLYOCFGRIPRELPPIHVEFN
 MPLEFLVONTIINSLSIAIIEKPTCPGICYSIETAPRETPPOSHLHLLPS
 TDGHLTLISGALIKNKHRTYVFRGSPMOFPRKMTHTASGRDLODACYGMSYD
 IKRIMEVYNPNLKYVSNITNPHKPDITNTIKTVAKRTDLVLSVLESTLPS
 CGKVMNITDSEKILPGETLNKLYTKYTKRNESTKAVMTTGTNTNPMVIEPTIW
 LPMTPLQVTKNPTNMTIIRKDLAIACVPYSTLEDROPAPSPSVFNPDLITIW
 EDSMNVATLGENIYSRCHLNKSENTPSPMDTP"
 complement(18848..19345)
 /note="ORF B04; short ORF of immediate early transcript 1
 RNA"
 /codon_start=1

Query Match 14.3%; Score 194.8; DB 14; Length 108873;
 Best Local Similarity 52.7%; Pred. No. 7.9e-37;
 Matches 520; Conservative 0; Mismatches 427; Indels 39; Gaps 3;

CDS
 RNA
 /codon_start=1

Query 316 atgagccagatattgacattatcagactctagagagtatgctaaagctgtctca 375
 Db 107637 ATAAACGAGACTGTGAGCGCTTCAGAGCAAGAGATTCAATACGTTCCACATGAGT 107696
 QY 376 aaggaaggaagaagctcccaatagccattcttggctgcacaaagaatgcaatag 435
 Db 107697 AAAGAAAGAGCTAGCTCCCATTTGCTCATGATGGTGGTGGATGAGAGATTGAAC 107756
 QY 436 gtggaaggtcttcagcttatatacaacagacaatttaccgacattatgat 495
 Db 107757 TTGAAAGGCTGCTGCAAGCGGTATAGCCCCCGAGACATATATCTGTGCGAGT 107816
 QY 496 cgttaagcacctatactctcaaatgtgcacatgaacatttagcttaagtctctccaat 555
 Db 107817 GTGAAGTCCCAAGACTTCAAGAGAGCGGTCAAGCGCATATTTGCTGCTTCCCAAT 107876
 QY 556 atttcatgtctccaataatagagctgtggaatatgcccacattccagaccceagct 615
 Db 107877 GTCTTCATGCGAGTAAAGTGTGTCGCGTGTATATGCTCCGTCACAGAGCAAGCT 107936
 QY 616 gatttaaatgtctgcagacctctgaagctctcaatccagtggaataatgttatcaac 675
 Db 107937 GACCTGAACGTGTATGAAAGCTTGCTCCAGAGTCAGTGCATGGAAGTACTTACTGAAAT 107996
 QY 676 ttgtgtgtgggaagaatttcccccggaagtcacaatttgaattgtgtgcagaagtgcgaataa 735
 Db 107997 ACAATGCGGAGAGACTTCCCATTAAGACCAATGCGAGATGCTTCGCGCCCAAGATG 108056
 QY 736 ctcaatggagaacaatatgttggagagcgtgtaaaccccccaacagtaaatgtgaagattc 795
 Db 108057 TTGAAGGCTTAAGAACACTATGAGAGTCTGAGTACTTCTGATAGCCAAAAAATTCGCTGG 108116
 QY 796 acttacacataatgaacttaagcgggtgcttatcgaataatgtgaagcttaacataagaa 855
 Db 108117 AAAATACCGCTATGAGTGAACAGACACTGTACCTT-----ACC 108155
 QY 856 aacatcccaagaagaagaccccccaataacatcacatatattgtgtgcagtgcttattt 915
 Db 108156 AGCAAGATGGAAGACCTCCCTCGATTAATTATTCACATGCTTACAGAGGAATGCTCTATTTT 108215
 QY 916 gtttaagtcagaagcattgttaaatatatttcaacaactccatcglttcaagcattttt 975

Db 108216 GTGAGCCTCTGAGACCTTTTCCAAATGCTTAGACAAACCTTAATCCAAAGACTGGTT 108275
 QY 976 gctgtctaaagaacataaacctctctgtatgcaacttttggctcaacttgaattggctt 1035
 Db 108276 GAATGGGTCAAGACACCTATAGCCCCGAGAACACCTCTGGGCCACCTTCAGGTGCT 108335
 QY 1036 ccaggaatacctcgtgggagattccagatcaagccagagatg--tgtcatctgcagagt 1092
 Db 108336 CCGTGATGCTCGTGGCTGCTGCTTCTGAGCCACCCCAAGTATACATCTCAGACATGACTGCC 108395
 QY 1093 aagactcgcctgtcaagtgtgaattactatgaagcttttc-----tat 1137
 Db 108396 ATCGCCAGGCTGCTCAAGTGGAGTACACAGGAGAGATGTACATGGGGGCGCTTAT 108455
 QY 1138 ccaggtgtgactggaactcactcgaagcgtgtattatgagctgcagaataag 1197
 Db 108456 GCACCTGCTCTGGAATCATGAGAGGCGCATCTGATTAGGGGGCGGAGCTGTAC 108515
 QY 1198 tggctatacaagaatggacatttggcttaataattgattcctaagtggaacctatc 1257
 Db 108516 TGGATCTCCAGAACCATCACCCTTGGCAACAAGTTGACCCGAGGTGATGATTAAC 108575
 QY 1258 ttgattaaatgcttgcagaagaagct 1283
 Db 108576 GTCTCTGACATGCTTAGAAGATGACT 108601

RESULT 6
 AB037597
 LOCUS
 DEFINITION
 beta-1,6-N-acetylglucosaminyltransferase A, complete cds.

AB037597 1206 bp mRNA ROD 20-OCT-2000
 Mus musculus Igfr A mRNA for
 beta-1,6-N-acetylglucosaminyltransferase A, complete cds.
 VERSION
 AB037597.1 GI:9650955
 KEYWORDS
 beta-1,6-N-acetylglucosaminyltransferase A,
 Mus musculus cell_line:F9 cell cDNA to mRNA.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 Chen,G.Y., Kurosawa,N. and Muramatsu,T.
 A novel variant form of murine beta-1,
 6-N-acetylglucosaminyltransferase forming branches in
 poly-N-acetylglucosamines
 Glycobiology 10 (10), 1001-1011 (2000)
 20485361
 JOURNAL
 MEDLINE
 2 (bases 1 to 1206)
 Kurosawa,N., Chin,G. and Muramatsu,T.
 Direct Submission
 Submitted (25-JAN-2000) to the DDBJ/EMBL/Genbank databases.
 Nobuyuki Kurosawa, Nagoya University School of Medicine, Department
 of Biochemistry, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550,
 Japan (E-mail: kurobetsuru.med.nagoya-u.ac.jp, Tel:81-52-744-2063,
 Fax:81-52-744-2065)

FEATURES
 source
 1..1206
 location/Qualifiers
 1..1206
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="F9 cell"
 1..1206
 /gene="Igfr A"
 1..1206
 /gene="Ignt A"
 /product="beta-1,6-N-acetylglucosaminyltransferase A"
 /codon_start=1
 /protein_id="BAB03496.1"
 /db_xref="GI:9650956"

translation="MPPSVRFVFFVSVTVTVIVLVISFGGDSQYOKINISDSVRL
 SQVSSPFDGSRFLMKRKMLHKEKPECTEVYVSHITPLPSDEPDLAVMYIH
 HNFDFALFRALFVNPONTVCVHDEKATAFKCAVQLVSCFPAALASKMEVYIG
 GISRLQADLNCIKDLSTSEVPMKAIINTCGDPLKTKNKIYVGLKGLKKNLTPGLV
 PPAHAIGRVYVREHREHLSKELSYVIRTTAKPPPHNLITVFGSAVVALSRFANFVLT

Fri May 3 10:57:51 2002

us-09-645-192-1.rge

Page 8

RDPRAYDLHWSKDTESPDEHFWTLNRPVPGSGMPNNAWGTGNLRAYKMMDEAKHH
GGCGHYVHGICLYGNGLDOWLINSQSLFANKFELNTYPLTVECELRLRERTLNGSE

BASE COUNT	300 a	318 c	285 g	303 t
ORIGIN				

Query Match	14.38;	Score 194.4;	DB 10;	Length 1206;
Best Local Similarity	52.18;	Pred. No. 8.3e-37;		
Matches 497; Conservative	0;	Mismatches 436;	Indels 21;	Gaps 2;

OY	337	tataagactcaaaaggattgtctcaaaaagctgtctcaaaaggaggaaagactccca	396
Db	223	TATGTGACCCAAACCAACCACTATATCACTGCCCCCTTTATCTCAGAGAAAGGTGCATTTTTCCT	282
OY	337	atagactatctcttggtgtctccaaaagatgcataatggttgaaaggtcatccatgct	456
Db	283	TTGGGTATGTCAATGATGCATATCCATCCAAATTTTGACACCTTTGCAAGGCTCTTCCAGGGCA	342
OY	457	atacaaacacgacacatattactatgcataccattatgcacgtgaagagacctatcacctc	516
Db	343	ATCTTCATGCTCTCAAAATATCTACTGTGTCTACGTGGATGAAAGGCAACAGCTCAATTC	402
OY	517	aaagttgcacatgaaacaatttagctaaagtgtctctccaaatttccattgcttcccaatia	576
Db	403	AAAGGTGCGCGGAGAACGTTTAGTAGAGTGCTGTTCCCAATGCTTCCTTGCTTCTTAAGATG	462
OY	577	gagcgtgtggaatatgtccacatttccagctccagcgtcatataaattgtctgtccgac	636
Db	453	GAGCGCGGTGCTATTGGTGGAAHTCCCGGCTCCAGGTGACTGAATGCTCAAAAGC	522
OY	637	ctctgtaagttccaatccagctgtggaatatgltatcaactgtgtgtgggaagaatttccc	696
Db	523	CTGTCCACCTCCGAGTGCCTCCCTGGAGTAGCGCATACAAACCGTGTGACAAAGACTTCCCC	582
OY	697	ctgaagtcacaattttagtaattgtgtgtcgaagtgtgaanaaacctaatgtgaagaatatgtgt	756
Db	563	CTTAAGAACCAACAAGGAGATAGTTCACTACCTTAAAGGGCTTAAGGGGAAGAACTCTACT	642
OY	757	gagaacgttgaaaccccccaaaacgtgaaattgaaagaattccactaccatcatgtaacttga	816
Db	643	CCGGGGGGTGCCTGCTCCAGGCCAACGCGCATTTGGAAGAGACAGGTAC-----GTCCACGG	696
OY	817	cgggtgccttatgtaattgtgaaagctcaaatgaagaacaacatctccagaagaagacacc	876
Db	697	GAACACCTTAACCAAGAGCTTCTTACGTATATGAACACAGCGCTTGAAGCCCCCACT	756
OY	877	ccccataacatcagataattggtgtgcagtgctatttgttttaagtcagaagattglt	936
Db	757	CCCCCAACCTCACCATTTATTTTGGCTCTGTTATGTGCCCCCTTCGAGAGAGTTTGCT	816
OY	937	aaatatatttcaacaacatccatcgtgttcaagacttlttgcctgtgtcctaagaacacac	996
Db	817	AACTTTGTCTCGTGACCCACGCGGCGTTGATTGCTCATTTGTCGCAAAATACTTTC	876
OY	997	tctccttgatgacaatttgggtctactctgattcgggtgtccagaatatccttggagatt	1056
Db	877	AGTCCCGATGTGCAATTTCTGGGTGACGCTCAATAGAGATTCCAGAGTCCCTGATCCATAG	936
OY	1057	tccagatccagcccgagatgtgtctgcagctgcagatgaagtcagcctgtctgtcgaaggaat	1116
Db	937	CCACCAACGATCTCTGACCGGTAACCTTAGAGCTGTGAAGCATGTGACATGGAAGG	996
OY	1117	tactatgaagctttctctatccagttgactcgtgactccctcgaagcgtgtgtat	1176
Db	997	AAGCATGTGAGGC-----TGCATATGTGCATCACTACCTCATGCAATTTGTATC	1041
OY	1177	tatgaagctgacaatatgaagtgtcctatcaagaatgacatgtggtgttctaatcaatt	1236
Db	1042	TATGGAACGGAAGACTTCGATGGCTGCTATTAATTCGAAAGCCTGTTTGCTAACAAATTT	1101
OY	1237	gattctaaagctggaacctatctgtattnaatgtcgttggcagaagaactgtgaaga	1290

Db 1102 GAACTCAACATACCCCTTACCGTGGAATGCCCTGGAAGGCTTCGAGAA 1155

[illegible]

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
1 (bases 1 to 1317)				
Schwentner, T. and Clausen, H.				
UDP-n-acetylglucosamine: galactose--g(b)1,3- n				
-acetylglucosamine--g(a)-r/(galactose to galactose)--g(b)1,6- n				
-acetylglucosaminyltransferases e, c29n63				
Patent: WO 0114535-A 14 01-MAR-2001.				
Schwentner, Tilo (DK); Clausen, Henrik (DK)				

```
source
1..1317
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1317
CDS
```

/code="The gene of C2GnT2"
 /nucleo_start=1
 /protein_id="CAC34676.1"
 /db_xref="GI:13366927"
 /translation="MVQKRRLCQLHLYMALCSTYMLAVAKLSFRLKCDSDHGLSES
 RESQYICRNLTYNFKLPAPKRSINCSQVFGDGEAVYQALINLDEVKKRPETPTD
 YLSLSTRDCEHHKAEKRFIQPLSKSEVEFPAYSMVIEKLEIKNERILRAYAAPOND
 CVHDEKSEPTFEKAVRAIISCIPNVFASLTVAYVASCRVQADLNCMDLLOSSS
 PWKFEINCGTGDPIKSAENAVQALKMLNGNSNSESVEPKHETFKMKEYHEVYRDT
 HLTKKKDPPRYNLMTIGNAYIASBPYQHYLNKPSQQLIEMVDTYSPDHLIAANA
 TLQKRRPAGSPVPHNPRTIDSDMTSIARLVMOGHEGIDGATYAPACSGIHQAICY
 YGADDLNMLQNHHLLANKPDKVDNADLOCLEEKRAIKAYGTEL"

BASE COUNT
ORIGIN

Query Match	14.1%;	Score 191.8;	DB 6;	Length 1317;
Best Local Similarity	52.5%;	Pred. No. 3.6e-36;		
Matches 517; Conservative	0;	Mismatches 432;	Indels 36;	Gaps 3;

Qy	317	tgaccacgatgatygtacattatcaagactccaagaagttatgctccaagaagctgtctcaa	376
Db	317	TCACCGAGAGACTGTGACACTTCAAGGCTGAAAGGAAAGTTTCATACAGTCCCACTGAGCA	376
Qy	377	aggaggaataaaagctctcccaatagcctatttttggttgtccacaagaatgcaatttgg	436
Db	377	AGGAAGAGGTGAGTGTCCCTATGTGATCTATAGGTGATTCAGAGAGATTGAAACT	436
Qy	437	ttgaaggcttatccatgctatatacaacagcascacatatttactgcatactatgac	496
Db	437	TTGAAAGGCTATGCGAGCTGTGATAGCCCTCAGAACATATAGTGTGCATGTGGATG	496
Qy	497	gtaaagcaccgtatcctctcaaaagttgcacatgacaactttgcctaagtgcttcccata	556
Db	497	AGAACTCCCAAAACTTTCAAAGGGGCTCAAGCAATATTTCTGTGCTCCCAATG	556
Qy	557	tttcatcgtcttccaattagaagctgtggaataagccacatttccagactccagcgtg	616
Db	557	TCTTATATGCCAGTAACTGCTGTGGGTGCTTATGCTCCTGCTCCAGGTCGAACGTG	616
Qy	617	attaaatgtctgttcggagccttctgaagttctcaatccagttggaatatgtatcaact	676
Db	617	ACCTCAACTGCATGAAACTTGTCTCCAGAGCTAGTCCCTGGAAATACCTTCGTGATA	676
Qy	677	tgttggtggcgaagattttcccttgaagttcaaatlttgaatttggttgtgtcagagtgtga	736
Db	677	CATGTGGAGCGACTTCTCTAATAAGACATGCAGAGATGTGCAGGCTCTCAAGATGT	736

QY 737 tcaatggaacaaatattgtgagacggtgaaacccccaacagtaatttggaaagattca 796
 Db 737 TGAATGGAGAAATAGCATGAGTCAGAGTACCTCTTACGACAAAGAAACCCGCTGGA 796
 QY 797 cttaaccatcagaacttgaacggtgacctatgaatatgttaaacctaccataagacaa 856
 Db 797 AATATCACTTTGAGGTAGTGAAGACACATTAC-----CTAACCA 838
 QY 857 acatctccaaggaagaccocccccaatacattcagatatgttggcagtgctatttg 916
 Db 859 ACAAGAAGAAGAGATCCCTCCCTTATATATTACATGTGTACAGGGAGATCCGACATG 898
 QY 917 tttaagtcagaagcatttgttaaatatattcaacaacatccatcggtcaagaacttttg 976
 Db 899 TGGCTTCCCGAATTTCTCCAAACATGTTTGAAGAACCTTAATCCCMACACTGATG 958
 QY 977 cctgtgtcaaaagacacatactctcctgattgagcaactttgggtaaccttgatcggttc 1036
 Db 959 AATGGGTAAAGACACTTAAGCCAGATGAACACTCTGGGCCACCTTACGCGTGCAC 1018
 QY 1037 caggaataacctgaggagatttccagatacagcccaagatg---tgtctgatttgaagat 1093
 Db 1019 GGTGATGCTGGCTCTCTTCCCAACCAACCCCAAGTACGACATCTCAACATGACTTCTA 1078
 QY 1094 agactcgccttgtaagtggaatttactatgaagccttttctatccagat----- 1143
 Db 1079 TTGCCAGGCTGGTCAAGTGGCAGGGTCTATGAGGAGACATCGATAGGGTCTCTTATG 1138
 QY 1144 ----tgtactggatctcactcacttgaagcggtgtgtatttatttgaagctgcagaatgaagt 1198
 Db 1139 CTCCCTGCTGTGATTCACACGCGGCTATCTCGTTTATGGGCTGGGACCTTGAAAT 1198
 QY 1199 ggcctatacaaatgagatgtgttgcataaatgttgcatttgaagtgagctact 1258
 Db 1199 GATCTCTCAACACCATCTCCTGTGGCCACACAGTTTGACCCCAAGGTAGATGATATG 1258
 QY 1259 tgattaatgcttggcagaanaagct 1283
 Db 1259 CTCCTCAGCTTAGAGATACCT 1283

RESULT 8

AF038650

LOCUS AF038650 1317 bp mRNA PRI 25-MAR-1999
 DEFINITION Homo sapiens core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase
 (C2/4GNT) mRNA, complete cds.
 ACCESSION AF038650
 VERSION AF038650.1 GI:4511860
 KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 1317)
 Schwiertek,T., Nomico,M., Levery,S.B., Merx,G., van Kessel,A.G.,
 Bennett,E.P., Hollingsworth,M.A. and Clausen,H.
 Control of O-glycan branch formation. Molecular cloning of human
 CDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase
 forming core 2 and core 4

TITLE

J. Biol. Chem. 274 (8), 4504-4512 (1999)
 99143102

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .1317
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1. .1317
 /gene="C2/4GNT"

CDS

1. .1317
 /gene="C2/4GNT"
 /note="core 2/4-Gnt"
 /codon_start=1
 /product="core 2/core 4
 beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="A02152.1"
 /db_xref="GI:4511861"
 /translation="MVQWRKLCQLHYLVALGCGYMLLATVALKLSFRLKQSDPHGLIES
 RESQSYCNIMILYNFLKLPARKSINCSGYTRGDQAVLQAILNINLEVKRREPFDTH
 YSLTRDCHEFKAEKRFLOEPLSKSEVEEPPIKXSVMIHEKIERFLRAYAPONIV
 CYHDEKSPETKREAVKATISCEPAVFLASKIVRYVAVASMSVQADLMQMEDLOSSV
 PMKYLNTGTDTPPIKSNEMQALKMLNGRSMSEVPKPKKEFRMYKHEFVYVDTL
 HLNRKRDPPPTNLMTFGNATIVASRDQVAVLNKPSQOLIEVAKDTISFDEHLMA
 TLQARWMPGSVNPHPKYDIDSMTISLARLVKMGEGDIDKAPAPCSGIHQRAICV
 YXAGDLMWMLQNHHLNLFDPKRVDMNLQCLEEYLRKAIYIGTEL"
 BASE COUNT 359 a 307 c 322 g 329 t
 ORIGIN

Query Match 14.1%; Score 191.8; DB 9; Length 1317;
 Best Local Similarity 52.5%; Pred. No. 3.6e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgattgtgacattatcacagacttaagaggttaattgttcaaaagctgtctcaa 376
 Db 317 TCACCAAGACTGTGACACTTCAAGGCTGAAGAGAAAGTTCATACAGTCCCACTAGCA 376
 QY 377 aggaagaaagagcttcccaatagcctattcttggltgtccaaaagaatgcaattaag 436
 Db 377 AAGAGAGGAGGAGTTCCTTATTCATCTGATGATTCATGAGAGATGAAGAACT 436
 QY 437 ttgaagagcttaccatgcctatatacaacagacacatatattctgcacatcatgac 496
 Db 437 TTGAAGGCTTACGCGAGCTGTGTATGCCCTAGANCAATATCTGTGTCATGTGATG 496
 QY 497 gtaagacactgataccttcaaaagtgtccatgaacaatttagctaaagtgtctccata 556
 Db 497 AGAAGTCCCGAAGAACTTCAAGAGCGGCTCAAGCAATATTCTTGCTTCCCAATG 556
 QY 557 ttccatgtcttccaaatagagagctgtgaaatagccacatttccagactcaagctg 616
 Db 557 TCTTCATAGCCAGTAACCTGCTTCCGCTGCTTATGCTCTCTGCTCCAGGCTCAAGCTG 616
 QY 617 atttaattgcttgcagaccttcgaagcttcaatccagltggaatatgtatcaact 676
 Db 617 ACTCAACTGCATGAGAGACTGTGCTCAGAGCTCAGTCCGCTGGAATACTTCTGAATA 676
 QY 677 tgtgtggcaagatttccctcgaagtcacaatttgaattggtgtcagagtgtaaaaac 736
 Db 677 CATGTGGACGAGACTTTCCTATTAAGAGCAATGACAGATGTCAGAGCTCTTAAGATGT 736
 QY 737 tcaatggaacaaatattgtgagacggtgaaacccccaacagtaaatgtgaagattca 796
 Db 737 TGAATGGAGAAATAGCATGAGTCAGAGTACCTCTTACGACAAAGAAACCCGCTGGA 796
 QY 797 cttaaccatcagaacttgaacggtgacctatgaatatgttaaacctaccataagacaa 856
 Db 797 AATATCACTTTGAGGTAGTGAAGACACATTAC-----CTAACCA 838
 QY 857 acatctccaaggaagaccocccccaatacattcagatatgttggcagtgctatttg 916
 Db 859 ACAAGAAGAAGAGATCCCTCCCTTATATATTACATGTGTACAGGGAGATCCGACATG 898
 QY 917 tttaagtcagaagcatttgttaaatatattcaacaacatccatcggtcaagaacttttg 976
 Db 899 TGGCTTCCCGAATTTCTCCAAACATGTTTGAAGAACCTTAATCCCAACTGATG 958
 QY 977 cctgtgtcaaaagacacatactctcctgattgagcaactttgggtaaccttgatcggttc 1036
 Db 959 AATGGGTAAAGACACTTAAGCCAGATGAACACTCTGGGCCACCTTACGCGTGCAC 1018
 QY 1037 caggaataacctgaggagatttccagatacagcccaagatg---tgtctgatttgaagat 1093

/note="glycosyltransferase; core 2 Gnt-II"
 /codon_start=1
 /product="beta-1,6-N-acetylglucosaminyltransferase"
 /protein_id="AA010824.1"
 /db_xref="GI:4204684"
 /translation="MVQWRRLQQLHYLMALGCMYLMATVALKLSFRKCDSDHGLIES
 RESOQYCRNLIYNFLKLPARKSINCSVTGQDQAVLQALINLEVKRREPDTIH
 YLSLRDCEHFAERKFIOPFLSKREVEPIAYSVIHEKIEFELLRAVYAPQNIY
 CVHDEKSEPEFEKAVKAIISCFPNVFIASKVRVYASMSRVOADLNCMEDLQSSV
 PKRYFLNTGCTDEPIKSNAMNOALKMLNGRSMSESVPPKHKEKRYHFEVYDPL
 HLTKRKDDPPYNIILMTFGNAYIVASRDYVQHLKPKKQQLIENVKQDYSDEHMA
 TLQRRKRMGSPVNHPRKIDISDMTSIARLVKMGHGDIDKGAPVPCSGIHQRAICV
 YGAGDLNMWLNQHNHLANKFDPKVDNALQCLEEYLRYKAIYGTGL"
 YGAGDLNMWLNQHNHLANKFDPKVDNALQCLEEYLRYKAIYGTGL"

BASE COUNT 569 a 477 c 526 g 556 t
 ORIGIN
 Query Match 14.1%; Score 191.8; DB 9; Length 2126;
 Best Local Similarity 52.5%; Pred. No. 3.6e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
 QY 317 tgacagtgatgtgacattatcagaactcaagaggctatgtccaaagctgtccaa 376
 DB 670 TCACGAGAGACTGTGAGACTTCAAGGCTGAAGAGATTCAATGATCCACTGACA 729
 QY 377 agggaggaagaagcttcccaatagcctattcttgggtgtccaaagaatgcaatttg 436
 DB 730 AAGAAGAGGTGAGTCCCTATTGCATACCTGTATGGATTCAGAAAGATTGAAACT 789
 QY 437 ttgaagagcttatccatgctatatacaaccagacaatatttaccatcatatgac 496
 DB 790 TTGAAGGCTACTGAGAGCTGTATGCGCCCTCAGAACATATGATGTCATGATGATG 849
 QY 497 gtaaggacactgataacttcaagattgcatgacaatttagctaagtgtctccata 556
 DB 850 AAGAATCCCGAAGAACTTTCAAGAGGGGCAAGCAATATTCTTCTGCTCCCAATG 909
 QY 557 ttctcatgtctccaatttagagctgtgataatgcccacatttccagagctcaggctg 616
 DB 910 TCTTCATGACCAAGTACGTGTGCGGTGCTTATGCTCTGCTGCTCAGAGGTGCAACTG 969
 QY 617 attaaattgcttgcgaacttcttgaagcttcaatccagtggaatattgttcaact 676
 DB 970 ACCCAACTGCATGGAAGACTTCTCCAGAGCTCAGTCCCTGGAATACTTCTCGATA 1029
 QY 677 tgttggcgaagatttccctcgtgaagcaatttgaattgttgcagagtggaataaac 736
 DB 1030 CATGTGGAGCGACTTCTTAAAGACAAATGCAGAGATGTCAGGCTCTCAAGATGT 1089
 QY 737 tcaatgagacaatatgttggagagcgtgaaaccccaaacagtaaatgtgaaagattca 796
 DB 1090 TGAATGGAGGAATAGCATGAGTACAGAGTACTCTTAACCAAGAAACCCGCTGGA 1149
 QY 797 cttaaccatcagaacttagaagcgtgctttagaattgtgaacttaccataaaggcaa 856
 DB 1150 AATATCACTTGTAGGTAGTGAAGACACATTTAC-----CTAACCA 1191
 QY 857 aacatccaaagaaagacccccataaactatcagataattgttggagtgattatttg 916
 DB 1192 ACAAGAGAGAGATCCCTCCCTTATATTTAATGATGTTACAGGGGAATCGTACATGTG 1251
 QY 917 tttaagcaagcaattttaaataatttcaacaatccatcgttccaagactttttg 976
 DB 1252 TGGCTTCCCGAAGATTTCCTCAACATGTTTGAAGAACCCCTAAATCCCAACATGTGAT 1311
 QY 977 cctgttcaagaaacacatactctcgtatgacacatttgggctacacttgaattcctggttc 1036
 DB 1312 AATGGTAAAGACACTATATGCGCAAGTGAACACTCTGGGCCACCTTACAGGTGAC 1371
 QY 1037 caggaaatcctggggagagttccagataagcccgagatg---tgtcgattcgaagatga 1093
 DB 1372 GGTGATGCTGCTGCTGCTGCTCCCAACACCCCAAGTACGACATCTGACATGACTTCTA 1431

QY 1094 agactgccttgcataagtggaattcagagagcttttctaccagt----- 1143
 DB 1432 TTCCCGAGCTGTGCAAGTGGCAGAGGTCATGAGACATCATAGAGGTGCTCTTATG 1491
 QY 1144 -----tgtactgattccacttccgaagcgtgttattatagagctgcaagattaagct 1198
 DB 1492 CTCCTGCTCTGGAACCCACAGCGGCTATCTGCTTATGCGGTGGGCTGGGACTGAAT 1551
 QY 1199 ggcttcaagaatgacattgttgcataaattgtattcctaagtgtagccctatct 1258
 DB 1552 GGATGCTTAAACCACTACCTGTTGGCCACAAGTTTACCCAAAGTAGTGAATAG 1611
 QY 1259 tgataaatgcttgacagaagact 1283
 DB 1612 CTCCTCAGTGTGAGAAATACT 1636

RESULT 11
 LOCUS AX045249 2216 bp DNA PAT 24-NOV-2000
 DEFINITION Sequence 1 from Patent WO0066727.
 ACCESSION AX045249
 VERSION AX045249.1 GI:11343799
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 2216)
 AUTHORS Adolf, G., Heider, K.H., and Sommergruber, W.
 TITLE Tumour-associated antigen
 JOURNAL Patent: WO 0066727-A 1 09-NOV-2000;
 Boehringer Ingelheim International GmbH (DE)
 Location/Qualifiers

FEATURES
 source
 1..2216
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..426
 427..1743
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC17334.1"
 /db_xref="GI:11343800"
 /translation="MVQWRRLQQLHYLMALGCMYLMATVALKLSFRKCDSDHGLIES
 RESOQYCRNLIYNFLKLPARKSINCSVTGQDQAVLQALINLEVKRREPDTIH
 YLSLRDCEHFAERKFIOPFLSKREVEPIAYSVIHEKIEFELLRAVYAPQNIY
 CVHDEKSEPEFEKAVKAIISCFPNVFIASKVRVYASMSRVOADLNCMEDLQSSV
 PKRYFLNTGCTDEPIKSNAMNOALKMLNGRSMSESVPPKHKEKRYHFEVYDPL
 HLTKRKDDPPYNIILMTFGNAYIVASRDYVQHLKPKKQQLIENVKQDYSDEHMA
 TLQRRKRMGSPVNHPRKIDISDMTSIARLVKMGHGDIDKGAPVPCSGIHQRAICV
 YGAGDLNMWLNQHNHLANKFDPKVDNALQCLEEYLRYKAIYGTGL"
 YGAGDLNMWLNQHNHLANKFDPKVDNALQCLEEYLRYKAIYGTGL"

BASE COUNT 600 a 499 c 547 g 570 t
 ORIGIN
 Query Match 14.1%; Score 191.8; DB 6; Length 2216;
 Best Local Similarity 52.5%; Pred. No. 3.6e-36;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;
 QY 317 tgacagtgatgtgacattatcagaactcaagaggctatgtccaaagctgtcctca 376
 DB 743 TCACGAGAGACTGTGAGACTTCAAGGCTGAAGAGATTCAATGATCCACTGACA 802
 QY 377 agggaggaagaagcttcccaatagcctattcttgggtgtccaaagaatgcaatttg 436
 DB 803 AAGAAGAGGTGAGTCCCTATTGCATACCTGTGATGTTCAATGAAAGATTGAAACT 862
 QY 437 ttgaagagcttatccatgctatatacaaccagacaatatttaccatcatatgac 496
 DB 863 TTGAAGGCTACTGAGAGCTGTATGCGCCCTCAGAACATATGATGTCATGATGATG 922
 QY 497 gtaaggacactgataacttcaagattgcatgacaatttagctaagtgtctccata 556

```

Db 923 AGAAGTCCCAAGAACTTCAAGAGGGGCTCAAGCAATATTCTTCTCCCAATG 982
Oy 557 ttcttgcctcccaatagaggtgtggaatagcccaattccagactcagagctg 616
Db 983 TCTTATAGCCAGTAAGTGTGTGGGTATTATGCTCTCTGCGCCAGGATGCAAGCTG 1042
Oy 617 atttaattgtctgtcgagcctctgaagcttcaatccagttggaataatgttcaact 676
Db 1043 ACCCTCAACTGATGAGAGCTTCTCCAGAGCTCAGTCCCTGGTGAATACTTCTGAATA 1102
Oy 677 tgtgtgggcaagatttcccccgaagccaatttgaattgtgtccagagtgaataaac 736
Db 1103 CATGTGGGAGGAGACTTCTCTATAAGAGCAATGACAGATGTGTCAGGCTTCAAGATGT 1162
Oy 737 tcaatggagcaaatatgttggagagcgtgaaaccccccaacagtaaatgtgaagatca 796
Db 1163 TGAATGGAGAGATGACATGAGTCAAGAGTACCTCTACACAAAGAAACCCGCTGGA 1222
Oy 797 cttaacatcatgaacttagacggtgcttatgaatatgtgaagctacaataagagaca 856
Db 1223 AATATCACTTTAGGTAGTAGAGAGACATTAAC-----CTAACCA 1264
Oy 857 acatctccaaggaagcaaccccccaataacatcagatatgttggcagtgctattttg 916
Db 1265 ACAGAGAGAGAGATCCTCCCTTAAATTTAACTATGTATTACAGGGAATCGTACATTTG 1324
Oy 917 tttaagtaagcatttgttaataatatttcaacaactccatgcttaagaactttttg 976
Db 1325 TGGCTTCCCGAGATTTCCGCAACATGTTTGAAGAACCCCTTAATCCCAACACTGATTTG 1384
Oy 977 ccttggtctaagaacacatactctcctgtatgagcatttgggtacctgtatcggttc 1036
Db 1385 AATGGGTAAAGACACTTATAGCCAGATGAACACCTCTGGGCGCACCTTCAGCTGCAC 1444
Oy 1037 caggaaatactgggagatttccagataagccccagatgt---tgcitacttgcagata 1093
Db 1445 GGTGATGCTTGCTGCTGTGTTCCCAACCAACCCCAAGTACGACTCTCAACATGACTTCTA 1504
Oy 1094 agactgcctgtcaagtgtgaattactatgaagccttttctatccagt----- 1143
Db 1505 TTGCCAGGCTGTCAAGTGGCAGGGTCAATGAGGAGACATCGATAGAGGTGCTCCTTATG 1564
Oy 1144 -----tgtaactgatactcccttgaagcgtgtgtatttattgagagctcagaataagt 1198
Db 1565 CTCCCTGCTGTGGAATCCACGCGGGCTATGCGTTTATGGGCTGGGACTTGGAAT 1624
Oy 1199 ggcctatcaaatgtgacattgttgcctaataatttgaattcgaagtggaccatct 1258
Db 1625 GGATGCTTCAAAACCATCACTGTTGGCCACAAAGTTGACCCAAAGTAGATGATATG 1684
Oy 1259 tgattaaatgtctggcagaaagct 1283
Db 1685 CTCTTCACTGCTTAGAAGATACCT 1709

RESULT 12
AC022480/C
ID AC022480 standard: DNA; HTG: 181983 BP.
XX AC022480;
AC AC022480;
SV AC022480.5
XX
XX
XX 06-FEB-2000 (Rel. 62, Created)
XX 18-JUL-2000 (Rel. 64, Last updated, Version 5)
DE Homo sapiens chromosome 15 clone RP11-361D15, WORKING DRAFT SEQUENCE, 24
DE unordered pieces.
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
OS

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
XX [1]
XX 1-181983
XX RA Waterston R.H.;
XX "The sequence of Homo sapiens clone";
XX Unpublished.
XX
XX [2]
XX 1-181983
XX RA Waterston R.H.;
XX Submitted (04-FEB-2000) to the EMBL/GenBank/DBJ databases.
XX Genome Sequencing Center, Washington University School of Medicine, 4444
XX Forest Park Parkway, St. Louis, MO 63108, USA
XX
XX On Jul 16, 2000 this sequence version replaced gi:7321974.
XX
XX Center: Washington University Genome Sequencing Center
XX Center code: WUGSC
XX Web site: http://genome.wustl.edu/gsc/index.shtml
XX
XX Center project name: H.NH0361D15
XX
XX Sequencing vector: M13; 100%
XX
XX Sequencing strategy: Dye-terminator Big Dye; 0% of reads
XX
XX Assembly program: Phrap; version 0.990319
XX
XX Consensus quality: 164956 bases at least Q40
XX
XX Consensus quality: 170052 bases at least Q30
XX
XX Insert size: 168000; agarose-fp
XX
XX Insert size: 179683; sum-of-ctnigs
XX
XX Quality coverage: 3.76 in Q20 bases; sum-of-ctnigs
XX
XX Quality coverage: 3.85 in Q20 bases; sum-of-ctnigs
XX
XX
XX * NOTE: This is a 'working draft' sequence. It currently
XX * consists of 24 contigs. The true order of the pieces
XX * is not known and their order in this sequence record is
XX * arbitrary. Gaps between the contigs are represented as
XX * runs of N, but the exact sizes of the gaps are unknown.
XX * This record will be updated with the finished sequence
XX * as soon as it is available and the accession number will
XX * be preserved.
XX
XX 1
XX 1087: contig of 1087 bp in length
XX
XX 1088
XX 1187: gap of unknown length
XX
XX 1188
XX 2741: contig of 1554 bp in length
XX
XX 2742
XX 2841: gap of unknown length
XX
XX 2842
XX 4520: contig of 1679 bp in length
XX
XX 4521
XX 4620: gap of unknown length
XX
XX 4621
XX 6998: contig of 2378 bp in length
XX
XX 6999
XX 7096: gap of unknown length
XX
XX 7099
XX 10122: contig of 3024 bp in length
XX
XX 10123
XX 10222: gap of unknown length
XX
XX 10223
XX 14393: contig of 4171 bp in length
XX
XX 14394
XX 14493: gap of unknown length
XX
XX 14494
XX 18603: contig of 4110 bp in length
XX
XX 18604
XX 18703: gap of unknown length
XX
XX 18704
XX 22356: contig of 3653 bp in length
XX
XX 22357
XX 22456: gap of unknown length
XX
XX 22457
XX 26948: contig of 4492 bp in length
XX
XX 26949
XX 27048: gap of unknown length
XX
XX 27049
XX 31773: contig of 4725 bp in length
XX
XX 31774
XX 31873: gap of unknown length
XX
XX 31874
XX 36373: contig of 4500 bp in length
XX
XX 36374
XX 36473: gap of unknown length
XX
XX 36474
XX 41562: contig of 5089 bp in length
XX
XX 41563
XX 41562: gap of unknown length
XX
XX 41563
XX 46645: contig of 4983 bp in length
XX
XX 46646
XX 46745: gap of unknown length
XX
XX 53335: contig of 6590 bp in length
XX

```



```

CC * 53336 53435: gap of unknown length
CC * 33436 60197: contig of 6762 bp in length
CC * 60198 60297: gap of unknown length
CC * 60298 68352: contig of 7955 bp in length
CC * 68353 75761: gap of unknown length
CC * 75762 75861: gap of unknown length
CC * 75862 84055: contig of 8194 bp in length
CC * 84056 84156: gap of unknown length
CC * 84156 94192: contig of 10037 bp in length
CC * 94193 94292: gap of unknown length
CC * 94293 109722: contig of 15430 bp in length
CC * 109723 125237: gap of unknown length
CC * 125238 125337: gap of unknown length
CC * 125338 142872: contig of 17535 bp in length
CC * 142873 159682: gap of unknown length
CC * 159683 159782: gap of unknown length
CC * 159783 181983: contig of 22201 bp in length.

XX *
FH
FH Key Location/Qualifiers
FH
FH source 1. .181983
FH /chromosome="15"
FH /db_xref="taxon:9606"
FH /organism="Homo sapiens"
FH /clone="Rp11-361D15"
FH
FH 1. .1087
FH
FH misc_feature /note="assembly_name:Contig19"
FH 1188. .2741
FH misc_feature /note="assembly_name:Contig24"
FH 2842. .4520
FH misc_feature /note="assembly_name:Contig25"
FH 4621. .6998
FH misc_feature /note="assembly_name:Contig26"
FH 7099. .10122
FH misc_feature /note="assembly_name:Contig27"
FH 10223. .14393
FH misc_feature /note="assembly_name:Contig28"
FH 14494. .18603
FH misc_feature /note="assembly_name:Contig29"
FH 18704. .22356
FH misc_feature /note="assembly_name:Contig30"
FH 22457. .26948
FH misc_feature /note="assembly_name:Contig31"
FH 27049. .31773
FH misc_feature /note="assembly_name:Contig32"
FH 31874. .36373
FH misc_feature /note="assembly_name:Contig33"
FH 36474. .41562
FH misc_feature /note="assembly_name:Contig34"
FH 41663. .46645
FH misc_feature /note="assembly_name:Contig35"
FH 46746. .53335
FH misc_feature /note="assembly_name:Contig36"
FH 53436. .60197
FH misc_feature /note="assembly_name:Contig37"
FH 60298. .68252
FH misc_feature /note="assembly_name:Contig38"
FH 68353. .75761
FH misc_feature /note="assembly_name:Contig39"
FH 75862. .84055
FH misc_feature /note="assembly_name:Contig40"
FH 84156. .94192
FH misc_feature /note="assembly_name:Contig41"
FH 94293. .109722
FH misc_feature /note="assembly_name:Contig42"
FH 109823. .125237
FH misc_feature /note="assembly_name:Contig43"
FH 125338. .142872
FH misc_feature /note="assembly_name:Contig44"
FH 142973. .159682
FH misc_feature

```

```

FT /note="assembly_name:Contig45 clone_end:SP6
FT vector_side:left
FT 159783. .181983
FT misc_feature /note="assembly_name:Contig46"
FT
SQ Sequence 181983 BP; 53047 A; 36575 C; 36896 G; 52718 T; 2347 other;

Query Match 14.18; Score 191.8; DB 33; Length 181983;
Best Local Similarity 52.5%; Pred. No. 4.3e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

Qy 317 tgacagtgatggaacttatcagactcgaaggttaagttatgccaagctgtgtccaa 376
Db 86723 TCACCAGGACTGTGAGACTTTCAGGCTGAAAGGAAGTTCATCACTTCCACTGACGA 86664
Qy 377 agagaggaagaagcttcccaatagcctatcttgggtgtccacaagaatgcaattg 436
Db 86663 AAGAAAGAGTGGAGTCCCTATGCTATGCTATGATGATGATGATGATGATGATGAAACT 86604
Qy 437 ttgaaggttatccatgtctatacaacagcacaatttatgtcatcatatgac 496
Db 86603 TTGAAGGCTACTGCGAGCTGTGATGCCCTCAGAACATATACTGTGCTCATGTGATG 86544
Qy 497 gtaagcacctgataccttccaagtgtgcatgaacaattgactaagtgcttcccaata 556
Db 86543 AGAAGTCCCGAAGAACTTTCAGAGGCGGTCAAGAGCAATATTTCTGTGCTCCCAATG 86484
Qy 557 tttaattgtcttcaaatatagagctgtggaatattgccaattccaagctcagctg 616
Db 86483 TCTTCATGCGAGTAAGTGTGTTGGGTGTTATGCTCTGCTGCTGCTGCTGCTGCTGCTG 86424
Qy 617 attaatgtctgtcgcagcttctgaagcttcaatccatccagtggaatatgttcaact 676
Db 86423 ACCTGATGCTGATGGAACACTTGTCTCCAGAGCTCAGTCCCTGGAATATCTCTGATA 86364
Qy 677 tgtgtggcaagatttccctcgtgaatgaatttgaattgtgtgcagagttgaaaaac 736
Db 86363 CATGCGGACGACGACTTCTCTTAAAGCAATGAGAGAGTGTGCTGCTGCTGCTGCTG 86304
Qy 737 tcaatggagcaaatatgttggagagctgtgaaccccaagaatgtgaagattca 796
Db 86303 TGAATGGAGGATATGCAATGAGTCAAGTCACTCTTCAACCAAGAAACCCGCTGGA 86244
Qy 797 cttaaccatcatgaacttgaagcgtgtcctatgaatatgtgaagctaaccaatgaagca 856
Db 86243 AATATCACTTGAAGTGTGAGAGACATATACAC-----CTAACCA 86202
Qy 857 acatctccaagaagcaccocccatacatgaatatgttggcagtgctatttg 916
Db 86201 ACAAGAGAGAGATCTCTCCCTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 86142
Qy 917 tttaagtcgaagcatttgaatatatttcaacaacatccatcgttcaaacattttg 976
Db 86141 TGGCTTCCCGAGATTTCTGTCACACATGTTTGAAGAACCTTAATCCACACATGATTG 86082
Qy 977 cctgtctcaagacacacacacacacacacacacacacacacacacacacacacacac 1036
Db 86081 AATGGGTAAAGACACTTATGACCCAGATGAACACTTGTGGCCACCTTGTAGCGTGAC 86022
Qy 1037 caggaataccttggggaagatttccagatcagccagagatg---tgtctatctgagagta 1093
Db 86022 GGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 85962
Qy 1099 agactcgccttgaagtggaatttactatgaagcttttctatccagct----- 1143
Db 85961 TTGCCAGGCTGTCAAGGCGAGGCTCATGAGGAGACATGATGAAGGCTCTCTTATG 85902
Qy 1144 -----tgtactgatacacttgaagcgtgtgtatctatggagcgtcaaatatggt 1198
Db 85901 CTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 85842
Qy 1199 ggcctatcaagaatgacatttggttgcctaataattgattcgaagtggaacctatct 1258

```

```

Db 85841 GGAGCTTCAAAACCACTGCTGTGGCCAAAGATTGACCAAGATGATGATGATG 85782
QY 1259 tgattaatgcttgccagaaaagct 1283
Db 85781 CTCTCACTGCTGTAGAAATGATCT 85757

RESULT 13
AC092755 187760 bp DNA HTG 27-JUL-2001
LOCUS AC092755
DEFINITION Homo sapiens chromosome 15 clone RP11-361D15 map 15q22, ***
SEQUENCING IN PROGRESS *** 4 ordered pieces.
AC092755 AC022460
AC092755.2 GI:15022677
HTG: HTGS_PHASE2; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 187760)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Pate, D. and Hood, L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
2 (bases 1 to 187760)
Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dors, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.
Pate, D. and Hood, L.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT
On Jul 27, 2001 this sequence version replaced GI:15021933.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMSC
Web site: http://chroma.mbl.washington.edu/msg_mw
Contact: leetowens@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Assembly program: Phrap; version 0.990399
Insert size: 185000; agarose-fp
Quality coverage: 10.4x in Q20 bases; sum-of-contigs
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 6055: contig of 6055 bp in length
* 6056 6155: gap of unknown length
* 6156 33287: contig of 27132 bp in length
* 33288 33388: gap of unknown length
* 33388 160326: contig of 126839 bp in length
* 160326 160327: gap of unknown length

```

```

FEATURES             * 160327 187760: contig of 27434 bp in length.
Source
  1..187760
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="15"
  /map="15q22"
  /clone="RP11-361D15"
  /clone_1db="RP11-361D15"
  /note="This clone overlaps RP11-112N19 and RP11-568G20"
BASE COUNT  53775 a 38839 c 38241 g 56604 t 301 others
ORIGIN
Query Match      14.1%  Score 191.8: DB 2: Length 187760;
Best Local Similarity 52.5%: Pred. No. 4.3e-36;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

QY 317 tgaccagtgatgtgacattatccagactcagaaggtatgtcgaagaagctgtctcaa 376
Db 40861 TCACGACAGACTGTGAGCACTTCAAGGCTGAAGAGCAAGTCAATACAGTCCACAGCA 40920
QY 377 agagagagaagaagctcccaatagctattcttgggtgtccacaagaatgcaattatg 436
Db 40921 AAGAAAGGTGAGTGCCTTATGCACTACTATGATGATGATGATGATGATGATGATG 40980
QY 437 ttgaagagcttaccatctatatacaaccagacacatattatctgacatctatgac 496
Db 40981 TTGAAAGGCTACTGCGAGCTGTGATGCGCCGACAGACATATAGTGTGATGATG 41040
QY 497 gtaagagcaccgtatcattcgaagtgtgcaatgcaatgcaatgcaatgcaatgcaat 556
Db 41041 AGAAGTCCCAAGAACTTTCAGAGGCGTCAAGCAATTTCTTCTTCTTCTTCTTCT 41100
QY 557 ttctcattgcttcccaatagagagctgtggaatagccacattccagactccagctg 616
Db 41101 TCTTCATACCGAGTACGCTGCTTGGGTGTTTATGCTTCCCTGCTGCTGCTGCTGCT 41160
QY 617 atttaattgcttgcagaccttcgaagctcctcgaatccacagtggaatgcaatgcaat 676
Db 41161 ACCTCACTGCAATGAGAGCACTTCTCAGAGCTCAGAGCTGCGGAGAAATCTCTGAA 41220
QY 677 tggtagagcaagatcttccctcgaagctcgaatcttgaattgtgcagagtggaataac 736
Db 41221 CATGTGGAGCACTTCTTCTTAAAGACATGACAGATGATGATGATGATGATGATG 41280
QY 737 tcaatgagacaatatgttgagagcgtggaacccccaacagtaattggaagatca 796
Db 41281 TGAATGGAGCAATACATAGCATGAGTCAAGGTACCTCTTACGACAAACCCCGTGA 41340
QY 797 ctgacatcatgacatagagcgtggtccttgaatatgtgagctacacataagaagaca 856
Db 41341 AATATACCTTGAAGTAGTAGAGACATTTAC-----CTTACCA 41382
QY 857 acatctcccaagagaagaccccccacaatcagatatgttgcagtggttattg 916
Db 41383 ACAAGAGAGAGATCTCCCTTATATTTATATTTATTTATTTATTTATTTATTTATTT 41442
QY 917 tttaagtgcaagatttgttaatatatttcaaacacccatcggttaagctttttg 976
Db 41443 TGGCTTCCGAGATTGTGTCACATGTTTGAAGACCTTAATTCACCAACATGATG 41502
QY 977 ccgagcctaagaacatctctccagtagagcactttgggtacactgtatcggttc 1036
Db 41503 AATGGGTAAAGACATTTATAGCCAGATGAAACACTGCGGCCACOCCTTACGCGTGAC 41562
QY 1037 caggaatccttggtggagatttccagatccagcagagatg---tgtctgctcgaagta 1093
Db 41563 GGTGATGCTGCTGCTGTTCACACACCCACCAAGATGACATCTCAGACATGACTTGA 41622
QY 1094 agactgccttgcaagtgaattactatgaagagcttttctatccagtt----- 1143
Db 41623 TTGCCAGGCTGCTCAAGTGGCAGGCTCATGAGGAGACATCATTAAGGTGCTCTCATG 41682

```


Fri May 3 10:57:51 2002

us-09-645-192-1.rge

Page 17

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 09:33:00 ; Search time 175.33 seconds
(without alignments)

6559.881 Million cell updates/sec

Title: US-09-645-192-1

Perfect score: 1362
Sequence: 1 atgaagatatcaaatgtta.....atctcaactaccaratcatga 1362

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq_1101:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	22	AA500045
2	503	36.9	781	22	AAH98678
3	191.8	14.1	1221	21	AA965675
4	191.8	14.1	1317	21	AA965669
5	191.8	14.1	2108	21	AA965670
6	191.8	14.1	2128	22	AA62134
7	191.8	14.1	2229	21	AA99109
8	191.8	14.1	2236	22	AAH34463
9	191.8	14.1	2319	21	AAA4623
10	185.2	13.6	1807	16	AA089201
11	185.2	13.6	1807	17	AAH16201

12	185.2	13.6	1807	19	AAV30006	Full length CDNA s
13	185.2	13.6	1807	19	AAV16000	Human beta-1,6-N-a
14	183	13.4	186	21	AAV19209	Human secreted pro
15	175.2	12.9	2105	15	AA061559	CDNA sequence of h
16	174.4	12.8	5010	19	AAH24042	Rat DHI CDNA. Rat
17	174.4	12.8	5010	21	AA065468	Diabetic rat heart
18	172	12.6	2204	19	AAH24043	Human core 2gnt DN
19	172	12.6	2204	21	AA065469	Human heart core 2
20	160.8	11.8	378	19	AAH80112	Coding sequence fo
21	121.2	8.9	2102	18	AAV30005	CDNA encoding a re
22	112.4	8.3	936	22	AAH58252	Oligonucleotide D1
23	112.4	8.3	936	22	AAH58253	Oligonucleotide D1
24	112.4	8.3	936	22	AAH58254	Oligonucleotide D1
25	112.4	8.3	936	22	AAH58255	Oligonucleotide D2
26	112.4	8.3	936	22	AAH58256	Oligonucleotide D2
27	112.4	8.3	936	22	AAH58257	Oligonucleotide D1
28	110.6	8.1	936	22	AAH58258	Oligonucleotide D1
29	110.6	8.1	936	22	AAH58259	Oligonucleotide D1
30	110.6	8.1	936	22	AAH58260	Oligonucleotide D1
31	110.6	8.1	936	22	AAH58261	Oligonucleotide D1
32	110.6	8.1	936	22	AAH58262	Oligonucleotide D2
33	110.6	8.1	936	22	AAH58263	Oligonucleotide D2
34	92.2	6.8	378	17	AAH16202	C2Gnt catalytic do
35	80.8	5.9	997	19	AAV59800	Human secreted pro
36	66.2	4.9	300	21	AAH1261	Human colon cancer
37	52.4	3.9	192	22	AAH52139	Murine beta-1,6-N-
38	49.4	3.6	777	19	AAV59681	Human secreted pro
39	47.4	3.5	307	21	AAH1854	Human secreted pro
40	46.2	3.4	6171	19	AAV52170	Streptococcus pneu
41	41.8	3.1	244	22	AAH58238	Oligonucleotide D1
42	40.6	3.0	244	22	AAH58239	Oligonucleotide D1
43	40.6	3.0	7430	19	AAH31250	E. coli j96 pathog
44	36.8	2.7	466	22	AAH138302	Probe #6980 used t
45	35.8	2.6	623	22	AAH12685	Human CDNA clone t

ALIGNMENTS

RESULT 1		
ID	AA500045	standard; DNA: 1362 BP.
XX	AA500045:	
AC	14-MAY-2001	(first entry)
XX		
DT		
XX		
DE		Human DNA encoding C2GNT3.
XX		
KW		Human; C2GNT3; Thymus-related disorder; cancer; tumour; adenoma;
KW		UDP-GlcNAc: Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase; sarcoma;
KW		malignant melanoma; breast cancer; cervical cancer; hypoplasia;
KW		hyperactivity; atrophy; thymus enlargement; autoimmunity; arthritis;
KW		leukemia; lymphoma; immunosuppression; AIDS; Wiskott-Aldrich syndrome;
KW		acquired immunodeficiency syndrome; sepsis; wound healing; infection; ds
OS		Homo sapiens.
XX		
XX		
Key		Location/Qualifiers
FT	1..1362	
FT	CDS	
FT		/*tag= a
FT		/product= "C2GNT3"
FT		/complement (114..136)
FT		/*tag= b
FT		/label= "Binding site for PCR primer TSHC100"
FT		1347..1362
FT		/*tag= c
FT		/label= "Binding site for PCR primer TSHC121"
XX		
XX		W0200114535-A2.
XX		01-MAR-2001.
XX		

PF 24-AUG-2000; 2000MO-DK00469.
 XX
 XX 24-AUG-1999; 99US-0150488.
 PR
 PA (SCHW/) SCHWIENTER T.
 PA (CLAV/) CLAUSEN H.
 XX
 XX Schlientek T, Clausen H;
 PI WPI: 2001-226615/23.
 XX P-PSDB: AAU00037.
 DR
 XX New C26nr3 polypeptides and nucleic acids encoding the polypeptides
 PT useful for treating conditions mediated by a C26nr3 polypeptide, e.g.,
 PT thymus-related disorders, cancers, tumours, immunosuppression
 XX
 XX Claim 5; Fig 1; 97pp; English.

CC The sequence encodes Human UDP-N-acetyl-glucosamine:galactose-4-epimerase
 CC 3-N-acetyl-galactosamine-4-epimerase (UDP-GlcNAc:Galbeta1,3GalNAc alpha-R beta1,6GlcNAc transferase or
 CC C26nr3). C26nr3 and nucleic acids encoding it are useful in the
 CC preparation of compositions for treating a conditions mediated by C26nr3,
 CC particularly a thymus-related disorder. C26nr3, nucleic acids
 CC encoding it and antibodies against it may also be used for in vitro
 CC purposes related to scientific research, DNA synthesis and manufacture of
 CC vectors. In the prognostic and diagnostic evaluation of conditions
 CC associated with altered expression or activity of C26nr3 or conditions
 CC requiring modulation of C26nr3, as well as in monitoring conditions by
 CC detecting and localising the DNA and protein. Disorders such as tumours
 CC (e.g. malignant melanoma, adenoma and sarcoma) and cancers (e.g. of the
 CC breast or cervix), hypothyroidism, hyperactivity, atrophy, enlargement of
 CC thymus, autoimmune arthritis, leukaemia, lymphomas, immunosuppression,
 CC acquired immunodeficiency syndrome (AIDS), Wiskott-Aldrich syndrome,
 CC sepsis, wound healing, acute and chronic infection, cell-mediated or
 CC humoral immunity, or TH1/TH2 imbalance, may be treated using these protein
 CC or nucleic acid. The antibodies may be used to screen potential
 CC therapeutic compounds to determine their effects on a conditions such as
 CC thymus-related disorder or cancer, to determine the level of C26nr3
 CC expression in cells genetically engineered to produce C26nr3, or to
 CC detect and quantify polypeptides in a sample to determine their role in a
 CC particular cellular events or pathological states and to diagnose and
 CC treat such pathological states.

CC Sequence 1362 BP; 410 A; 260 C; 279 G; 413 T; 0 other;

Query Match 100.0%; Score 1362; DB 22; Length 1362;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaagatatcaaatgtattttaaaccacacacagcagaagaatttcacgtttt 60
 Db 1 atgaagatatcaaatgtattttaaaccacacacagcagaagaatttcacgtttt 60
 QY 61 ttaacctatgctcctccttctgttaagctcttaaatgtgagcagcctcttcgcaa 120
 Db 61 ttaacctatgctcctccttctgttaagctcttaaatgtgagcagcctcttcgcaa 120
 QY 121 aaagacattactgtgttagtacctcccaagtaacctgcctctgttaagaagaatcac 180
 Db 121 aaagacattactgtgttagtacctcccaagtaacctgcctctgttaagaagaatcac 180
 QY 181 actcatgttaagatgaagcagatgaagtaacgttcgttcgggtatcatgaacagag 240
 Db 181 actcatgttaagatgaagcagatgaagtaacgttcgttcgggtatcatgaacagag 240
 QY 241 ccttggaaattggaagaagctctggaataaagaagaagagacatattgacttggagat 300
 Db 241 ccttggaaattggaagaagctctggaataaagaagaagagacatattgacttggagat 300
 QY 301 gatgatgttctgtgcaatgaccagtgatgtgacattatcagaccttaagaaggttatgct 360
 Db 301 gatgatgttctgtgcaatgaccagtgatgtgacattatcagaccttaagaaggttatgct 360

Db 301 gatgatgttctgtgcaatgaccagtgatgtgacattatcagaccttaagaaggttatgct 360
 QY 361 caaagcttgcctcaaaagagagaaaagctccccaatgacctattcttggttgcac 420
 Db 361 caaagcttgcctcaaaagagagaaaagctccccaatgacctattcttggttgcac 420
 QY 421 aaagatcaattatggttgaagagcttatccatgctctatatacaaccagcaatattac 480
 Db 421 aaagatcaattatggttgaagagcttatccatgctctatatacaaccagcaatattac 480
 QY 481 tgcatacatatgacgtgaaagcagcctgataccttcaaaagtgcataaagaaattagct 540
 Db 481 tgcatacatatgacgtgaaagcagcctgataccttcaaaagtgcataaagaaattagct 540
 QY 541 aagtgcttcccaatttccatgtgcttcccaattgagagcttggaaattgacccacatt 600
 Db 541 aagtgcttcccaatttccatgtgcttcccaattgagagcttggaaattgacccacatt 600
 QY 601 tccagactcagagctgatttaaatgtctgtcgaaccttctgaagcttccatccagttg 660
 Db 601 tccagactcagagctgatttaaatgtctgtcgaaccttctgaagcttccatccagttg 660
 QY 661 aaatgttcaactgtgtggaagaatttccctcctgagcagcttgaatttgatg 720
 Db 661 aaatgttcaactgtgtggaagaatttccctcctgagcagcttgaatttgatg 720
 QY 721 tcaagattgaaaacaaactcaatgagcaaatatgttgagacggttgaacccccaaacgt 780
 Db 721 tcaagattgaaaacaaactcaatgagcaaatatgttgagacggttgaacccccaaacgt 780
 QY 781 aaatggaagatcattcatttccatcagaacttagaagcgttgcattgaaatgtgaaag 840
 Db 781 aaatggaagatcattcatttccatcagaacttagaagcgttgcattgaaatgtgaaag 840
 QY 841 ctacccaataagagacaacacatcccaagagagcagccccccaataacattcagatattgt 900
 Db 841 ctacccaataagagacaacacatcccaagagagcagccccccaataacattcagatattgt 900
 QY 901 ggcagctctatttctgttgaagcagcattgttaaatatttcaaacattccac 960
 Db 901 ggcagctctatttctgttgaagcagcattgttaaatatttcaaacattccac 960
 QY 961 gtcaagacatttctgcctgtctcaagacacatacctctcgtatgagcactttggct 1020
 Db 961 gtcaagacatttctgcctgtctcaagacacatacctctcgtatgagcactttggct 1020
 QY 1021 acccttgatcgggttccaggaatacctgggagagatttccagatcagccagagatgtgct 1080
 Db 1021 acccttgatcgggttccaggaatacctgggagagatttccagatcagccagagatgtgct 1080
 QY 1081 gatctgagagtaagactgcctgtctcaagtgaattatcatgaagctttctatccc 1140
 Db 1081 gatctgagagtaagactgcctgtctcaagtgaattatcatgaagctttctatccc 1140
 QY 1141 agttgacgtgatacctccttcgaagcgtgtgatttatgagcgtcagaattaaagtgg 1200
 Db 1141 agttgacgtgatacctccttcgaagcgtgtgatttatgagcgtcagaattaaagtgg 1200
 QY 1201 cttaacaagaatgagactgtgtgtctaatattgaattcctaagttgagccctatctt 1260
 Db 1201 cttaacaagaatgagactgtgtgtctaatattgaattcctaagttgagccctatctt 1260
 QY 1261 attaaatgcttgagcagaagaagcttgaagaacagcagagagacgttgatcacttgcctca 1320
 Db 1261 attaaatgcttgagcagaagaagcttgaagaacagcagagagacgttgatcacttgcctca 1320
 QY 1321 gaaagcttattatgagtagaataatctacacacacatcaga 1362
 Db 1321 gaaagcttattatgagtagaataatctacacacacatcaga 1362

RESULT 2
 AAH98678

ID AA98678 standard; cDNA; 781 BP.
 XX
 AC AA98678;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 535.
 XX
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 FN MO20154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001MO-US02657.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-061746.
 PR 03-AUG-2000; 2000US-0631451.
 PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HXSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX
 DR WPI: 2001-476164/51.
 XX
 PT P-FSDS; AAM24019.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1; Page 549; 1275bp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 SO Sequence 781 BP; 218 A; 175 C; 174 G; 214 T; 0 other;

Query Match 36.9%; Score 503; DB 22; Length 781;
 Best Local Similarity 96.3%; Pred. No. 1,4e-131;
 Matches 515; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 1 atgaagatattcaaatgtattatccaacataccctacagcagaagtttcatccttt 60
 DB 230 atgaagatattcaaatgtattatccaacataccctacagcagaagtttcatccttt 289
 OY 61 ttaaccctatggctgctctctttttaagcttcaatgtaagcagactttccgcaa 120
 DB 230 ttaaccctatggctgctctctttttaagcttcaatgtaagcagactttccgcaa 349
 OY 121 aaagacattacttggttgagctactccctaagctccttcgcttttgaagaacagat 180
 DB 350 aaagacattacttggttgagctactccctaagctccttcgcttttgaagaacagat 409
 OY 121 actcgttaagatgaagatcaggtatgaatgaatgaatggttggttactatgaacagag 240
 DB 410 actcgttaagatgaagatcaggtatgaatgaatgaatggttggttactatgaacagag 469
 OY 241 cctttggaatgtgaagaagctctggaataagaagaagagcaccatgacttgagagat 300

DB 470 cctttggaatgtgaagaagctctggaataagaagaagagcaccatgacttgagagat 529
 OY 301 gatgatgttggtgcaatgacacagatgttgacattatcagacttaagaagttatgct 360
 DB 530 gatgatgttggtgcaatgacacagatgttgacattatcagacttaagaagttatgct 589
 OY 361 caaagcttgctcctaagagaggaagaagcttcccatagcactatcttggttgccac 420
 DB 590 taaagcttgctcctaagagaggaagaagcttcccatagcactatcttggttgccac 649
 OY 421 aaagatgaattatggttgaaggcttcatcgtctatatcaacagacatatctac 480
 DB 650 caagaagcaattatggttgagagcttcatcgtctatatcaacagacatatctac 709
 OY 481 tgcatcattatgatcgttaagcagcttgatccttcaaaagttgcctgaacaatt 535
 DB 710 tgcatcattatgatcgttgagcggggagcacttggaaccttaaaagttgcctgaacaatt 764

RESULT 3
 ID AAA96575
 AA96575 standard; DNA; 1221 BP.
 XX
 AC AAA96575;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Core 2 beta-1,6-N-acetylglucosaminyltransferase catalytic region DNA.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase: core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis; ss.
 XX
 OS Homo sapiens.
 XX
 PN CA2296936-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 PR 03-FEB-1999; 99US-0118674.
 XX
 PA (GLYC-) GLYCODESIGN INC.
 XX
 PI Korczak B, Lew A;
 XX
 DR WPI: 2000-594746/57.
 XX
 PT New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 XX
 PS Claim 4; Page 54; 66pp; English.

The present sequence encodes a fragment of a human core 2
 beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 The polypeptide can be used to treat diseases and disorders, such as
 cancer, cardiovascular disorders and inflammatory disorders including
 asthma, rheumatoid arthritis, inflammatory bowel disease,
 arteriosclerosis, septic shock, adult respiratory distress syndrome
 (ARDS) and cancer. Various platelet-mediated pathologies such as
 atherosclerosis and clotting can also be treated. The polypeptides of
 the invention are predominantly expressed in gastrointestinal tissue
 (stomach, colon, intestine, testis) and are elevated in cancer.
 Gastrointestinal disorders that may be prevented or treated include
 ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 and ulcerative colitis. The antibodies may be used in

CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

XX
 SO Sequence 1221 BP; 342 A; 285 C; 295 G; 299 T; 0 other;

Query Match 14.1%; Score 191.8; DB 21; Length 1221;
 Best Local Similarity 52.5%; Pred. No. 5,7e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgattgacattatcagactcgaaggtatgctcaaaactgtctcaa 376
DB 221 tcaccagagactctgagacactcaaggctgaaggaaattcacaacattccactgagca 280
QY 377 agagaggaagaaagcttcccaataagccattcttctgtgtcccaagaatgcaattatg 436
DB 281 aagaagaggttgaggttccctatctcattgcatctatgcatcattgagaattgaaaact 340
QY 437 ttgaagagctatccatcagctataaacaacagacacatattctcgtacatccattatgac 436
DB 341 ttgaagagctatctgagagctgtaatagccccccaagaacatatactgtctcatgtgagt 400
QY 497 gtaaggaacctgataccctcaaaagtgcatagaacaattagctaaagtctctccaata 556
DB 401 agaagtcocccagaaccttccaagagcggtccaagacattattctgtcttcccaatg 460
QY 557 ttctcatgtctccaatttagagctgtgtaatatgcccacattccagactccaagctg 616
DB 461 tctccaagccaagtaagctgtctcggtgttattgtcttccctcggtccaaggggtcgaagctg 520
QY 617 attaaattgcttgcgcagcttctggaagcttccatccagctggaataatgataccaact 676
DB 521 acctcaactgcatggaagactgtgtccagagctccagagctcagtggaattacttccgaata 580
QY 677 tgtgtggcaagatttccctcgtaagtccaatttgtaattgtgtcagagttaaaaaaac 736
DB 581 catgtggagaggaacttccctcaataaagagcaatgcaagctgctccaagctctcaaatgt 640
QY 737 tcaatggagcaaatatgttgagacaggtgaaacccccaacagtaaatggaagaattca 736
DB 641 tgaatcgtggaggaatagatagtagagcagaggtactccttaagcacaagaacccgcgtgga 700
QY 797 cttaacatcatgaacttaagacgggtgctcttaagaattgtgaagctacccaataaggaca 836
DB 701 aatatcaacttgagtagtagtagagacacattaac-----ctaacca 742
QY 857 acatctccaaggaagacacccccccaataacattcagatattgttgcagtgcttatttg 916
DB 743 acaagaagaagagatcccccctataataactaatttacaaggaatgctgacattg 802
QY 917 tttaagtcagaagcattgttaaatatatttcaacaactccatcgttcaagactttttg 976
DB 803 tggcttcccgagattctgtccaacatgttttgaagaccctaatacccaacaactggtt 862
QY 977 ccctgtctaaagacacatactctctctatagacacttttgagcttacttctgactggttc 1036
DB 863 aatgggcaaaagacacttaagcccaatgaacccctctgagccaccccttcagcgtgac 922
QY 1037 caaggaatacctgggagagattccacagatcagcccaagatg---tgtctgatatcagaatg 1093
DB 923 ggttgatgcctgctctgttctccacaacacccaagtagcagacatcagacatgacttcta 982
QY 1094 agactgcctgttcaagctggaattactatgaagcttcttctatccacagf----- 1143
DB 983 ttgcacagctggttcaagttgcagaggtcactaagggagacatcagataaggggtcctctatg 1042
QY 1144 -----tgtactgatactcacccttgcagacgctgtgtattatagagctgcagaaatgaagt 1198
DB 1043 ctccctctcgtgaatccacacagcggtctatctgtttatggggctgggagacttgaatt 1102
QY 1199 ggctatacaagatggaactgtgttgcataataatttgattcagaagtaccctatct 1258
  
```

```

DB 1103 ggaatgcttaaaaccatccactgttgcgaacaagttgacccaaggtagatgataatg 1162
QY 1259 tgattaagcttgcagagaagct 1283
DB 1163 cctctcagtgcttaagaagaataact 1187
  
```

```

RESULT 4
AAAG6569
ID AAA96569 standard; DNA; 1317 BP.
XX
AC AAA96569;
XX
DT 08-FEB-2001 (first entry)
XX
DE DNA encoding a core 2 beta-1,6-N-acetylglucosaminyltransferase.
XX
KW Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
KW platelet-mediated pathology; arteriosclerosis; gastrointestinal disorder;
KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
KW diverticulitis; ulcerative colitis; ss.
XX
OS Homo sapiens.
XX
FH Key 1.1317 Location/Qualifiers
FT CDS
FT /tag= a
FT /product= "core 2
FT beta-1,6-N-acetylglucosaminyltransferase"
PN CA229636-A1.
XX
PD 03-AUG-2000.
XX
PF 03-FEB-2000; 2000CA-2296936.
XX
PR 03-FEB-1999; 99US-0118674.
XX
PA (GLYC-) GLYCODESIGN INC.
XX
PI Korczak B, Lew A;
XX
DR WPI; 2000-594746/57.
XX
DR P-PSDB; AAB18995.
XX
PT New nucleic acid molecules of core 2
PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
PT compositions for treatment of disorders mediated by the enzyme
PT including cancer, cardiovascular and inflammatory disorders.
XX
PS Claim 4; Page 50; 66pp; English.
XX
  
```

The present sequence encodes a human core 2
 beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 The polypeptide can be used to treat diseases and disorders, such as
 cancer, cardiovascular disorders and inflammatory disorders including
 asthma, rheumatoid arthritis, inflammatory bowel disease,
 arteriosclerosis, septic shock, adult respiratory distress syndrome
 (ARDS) and cancer. Various platelet-mediated pathologies such as
 atherosclerosis and clotting can also be treated. The polypeptides of
 the invention are predominantly expressed in gastrointestinal tissue
 (stomach, colon, intestine, testis) and are elevated in cancer.
 Gastrointestinal disorders that may be prevented or treated include
 ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.

Sequence 1317 BP; 359 A; 307 C; 322 G; 329 T; 0 other;

Query Match 14.1% Score 191.8; DB 21; Length 1317;
 Best Local Similarity 52.5%; Pred. NO. 5.9e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgatgtgacattatatacagacttaagaggltatgtctcaaaagctgtctcaa 376
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 317 tcaccagagactgtgagacttcacagctgaaggaagttcatcagcttcccactgagca 376
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 377 agggaggaaaaagcttcccaatagcctattcttgggtgtgtcccaaaagatgcaatttg 436
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 377 aagaagagggtgaggttccctatgtactctatgtatgtatctatcagaaagatgaaact 436
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 437 ttgaaaggttatccatgtctatatacaaccagcaaatatttactgtcaatcatatgac 496
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 437 ttgaaaggtactgcgagctgtgtatgtccctcagaaacatactatgtgtcatgtgagt 496
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 497 gtaagggaccctgataccttcaagtgtgcacagaaacttgcgaagtgtctccacata 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 497 agaagtcctccgaanaacttcaagaggcggtcaagaacatcattctgtcttcccaatg 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 557 ttctatgtctcccaatagaagctgtgagataatgtcccaatctccagacttccagctg 616
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 557 tcttcatagccagtaagctgtgttcgggtgtgttatgctctctgttccaggggtgcaagctg 616
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 617 atttaaatgtctgtcggagccttctgaagcttcaatccagttggaataatgttcaact 676
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 617 acctcaactgtcatggaagaaactgtctccagagctcagtgccgttggaanaacttccatga 676
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 677 ttgtgtggcgaagatttccctcgaagtcacaaatttgaattgtgtcagagttgaaaaaac 736
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 677 catgtgtggagcagacttctctataaagagacatgtgcagagatgttccagcttctcaagagt 736
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 737 tcaatgtggagcaaatctgtgtgagagcgtgtgaaccccccaacagtaattgtgaagatca 796
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 737 tgaatgtggagagatagcatgtgagtcagagatcaccctcaagacaaagaaacccgctgga 796
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 797 cttaaccatcatgaacttgaagcgggtgtcttgaatatgtgaagctacataaagaa 856
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 797 aatcatcactttaggttagtgagagacacataaac-----ctaaacca 838
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 857 acatctccaaggaagaccoccccaataacatcagatatgttggcagttcattatttg 916
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 839 acaagaagaagatccctccoccttataatctaactatgtttaacagggatcgtacatgt 898
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 917 tttaagcaaacatttgtttaatatatttcaaacactcatgttcaagaactttttg 976
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 899 tggcttcccgagatttctgtccaaactgttttgaagaacccttaaccccaacacatgactg 958
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 977 cctgtctaaagacacataactctcctgtatgtgacacttttgggtactgttatcgggttc 1036
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 959 aatgggttaaaagacactatagccagatgaacacctctgtggccaccccttcagcgtagac 1018
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1037 cagggatcactgtggagagatttccagatcagcccaagatg---tgtctatctgcagata 1093
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1019 ggtgtgatgcgtgcgtctgttcccaaccaccccaagtaacgaatctccagacatgacttca 1078
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1094 agactcgccttgcagaagtgaattactatgaaggttcttctcatccaggt----- 1143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1079 ttgccaggtgtgtcaagtgtgcaggggtcatgtgagggagacatcgatgaaggtgtcctctatg 1138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1144 -----tgtactgtatctaaccttgcgaacggtgtgtattatgagagctgcagaatlaagt 1198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1139 ctccctgtctgtgaatcccaacgaggtcatctgttattatgtgggtgtggagactgtaatt 1198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1199 ggttatcaaaagatgacactgtgtgtcttaataatattgtatcctaagttgacacttact 1258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1199 ggaigtcttaaaaacacataccctgtgtgcacaaagtttgacccaaggttagatgttaagt 1258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1255 tgatttaaatgtcttgccagaanaagct 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1259 ctcttcagtgcttagaagaataact 1283
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 5
 ID AAA96570
 AA AAA96570 standard; DNA; 2108 BP.

AC AAA96570;
 DT 08-FEB-2001 (first entry)
 XX

A core 2 beta-1,6-N-acetylglucosaminyltransferase DNA fragment.
 XX
 XX Core 2 beta-1,6-N-acetylglucosaminyltransferase; core2b GlcNAc-T;
 KW cancer; cardiovascular disorder; inflammatory disorder; asthma;
 KW rheumatoid arthritis; inflammatory bowel disease; arteriosclerosis;
 KW septic shock; adult respiratory distress syndrome; ARDS; cancer;
 KW platelet-mediated pathology; atherosclerosis; gastrointestinal disorder;
 KW clotting; ascites; cholelithiasis; cirrhosis; Crohn's disease;
 KW diverticulitis; ulcerative colitis; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX CA2296936-A1.
 PN
 XX 03-AUG-2000.
 PD
 XX
 PF 03-FEB-2000; 2000CA-2296936.
 XX
 XX 03-FEB-1999; 9905-0118674.
 PR
 XX (GLYC-) GLYCODESIGN INC.
 PA
 XX Koczak B, Lew A;
 PI
 XX MPI; 2000-594746/57.
 DR
 XX
 XX New nucleic acid molecules of core 2
 PT beta-1,6-N-acetylglucosaminyltransferase useful for providing new
 PT compositions for treatment of disorders mediated by the enzyme
 PT including cancer, cardiovascular and inflammatory disorders.
 PT
 XX
 XX Claim 4; Page 51-52; 66pp; English.
 PS
 XX
 XX The present sequence encodes a partial human core 2
 CC beta-1,6-N-acetylglucosaminyltransferase (core2b GlcNAc-T) polypeptide.
 CC The polypeptide can be used to treat diseases and disorders, such as
 CC cancer, cardiovascular disorders and inflammatory disorders including
 CC asthma, rheumatoid arthritis, inflammatory bowel disease,
 CC arteriosclerosis, septic shock, adult respiratory distress syndrome
 CC (ARDS) and cancer. Various platelet-mediated pathologies such as
 CC atherosclerosis and clotting can also be treated. The polypeptides of
 CC the invention are predominantly expressed in gastrointestinal tissue
 CC (stomach, colon, intestine, testis) and are elevated in cancer.
 CC Gastrointestinal disorders that may be prevented or treated include
 CC ascites, cholelithiasis, cirrhosis, Crohn's disease, diverticulitis
 CC and ulcerative colitis. The antibodies may be used in
 CC immuno-histochemical analysis, to detect the novel polypeptide and to
 CC localize it to particular cells and tissues and to specific subcellular
 CC locations and to quantitate the level of expression.
 CC
 XX
 XX Sequence 2108 BP; 596 A; 464 C; 509 G; 539 T; 0 other;

Query Match 14.1% Score 191.8; DB 21; Length 2108;
 Best Local Similarity 52.5%; Pred. NO. 7.2e-44;
 Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgaccagtgatgtgacattatcaagacttaagaggltatgtctcaaaagctgtctcaa 376
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 584 tcaccagagactgtgagcaacttcaagcgtgaaaggaagttcatcagcttcccactgagca 643
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 377 agggaggaaaaagcttcccaatagcctattcttgggtgtgtcccaaaagatgcaattatg 436
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 644 aagaagaggtgagttccctatgcatcactatggttcatcatgagaagaattgaaaaact 703
QY 437 ttgaaagccttaccatgcttatatacaacccagacataatttaccgcatcatatgctc 496
Db 704 ttgaaagccttaccatgcttatatacaacccagacataatttaccgcatcatatgctc 763
QY 497 gtaaggaaccttataacctcaaaagttgcaatgaacaatttgaatgcttcttccaaata 556
Db 764 agaagtcctcccaagaaccttcaaaagagcggtcaaaagaacatttcttcttcccaaatg 823
QY 557 ttcttaagcttcccaaatgagagctgtggaatagccacatttccgactccagagctg 616
Db 824 tcttcaatagcagtaagctgtgttcggttcaatgacctcccgagcggtcgcaagctg 883
QY 617 atttaattgcttgcgaaccttctgaagcttcaatccactggaatgaatattgttcaact 676
Db 884 acctcaactgcataggaagaccttgcctccagagctcagtgccgtggaataacttccctgaatg 943
QY 677 tgtgtgggcaagaatttccctcgaagtcgaatatttgaattgtgtcagagtgtgaaaaaac 736
Db 944 catgtgggagcgaaccttctctataaagaacatgacagatggttccagcttccaaagatgt 1003
QY 737 tcaattggagcaaatatgttggagacggtgaaacccccaacagtaaatgtgaaagatgtca 796
Db 1004 tgaattggagagatagcatgagtgagtcagaggtactccttaagacaaagaacccgctgga 1063
QY 797 cttaacatcatgaacttgaagcaggtgaccttgaatattgtgaagctaccataagaacaa 856
Db 1064 aatatacactttaggttagtgagaagaacattcac-----ctaaaca 1105
QY 857 aatctccaaaggaacaccccccaatacatcagataattgttgcagtgcttattgtt 916
Db 1106 acaagaagaagagatccctcccttataatttactaactgtttacaggagatgcgtacatgt 1165
QY 917 tttaagtaagcatttgttaatatatttcaacaactccatgcttcaagaactttttg 976
Db 1166 tgggtcccggaatttcgctcaacacatgttttgaagaaccttcaatcccaacactgattg 1225
QY 977 cctgtgctaagaacacatactctccttgatgagacatttgggctccttgatcgaggttc 1036
Db 1226 aatggtaaaagacacttatacagatgaacacactctgggcacaccttcaagcgtgcac 1285
QY 1037 cagggaatacttgggagatattccagatccagcccaagag--tgcctatctgcagata 1093
Db 1286 ggttgatgctgtcctgttctccaaacaccccaaglaagacatcccaacatgacttcta 1345
QY 1094 agactgccttgcagtggaattactaagaagcttttctatccagct----- 1143
Db 1346 ttgcccaggtgtgtaagtgaggggtcatgagggagacatgataagggtgctcctatgt 1405
QY 1144 -----tgtactgagctcacccttgcgaagcggtgtattatttgaagctgcagatgaagt 1198
Db 1406 cctcctgtcctggaatcccccagcggtgctatctgcttatttgggttgggcttgaatt 1465
QY 1199 ggcattacaagaatgagacattgttgcataaataatgtatcctaaggtgagccctact 1258
Db 1466 ggaatgcttcaaaacatcaacccctgttgcacaagaagttgacccaaggttagtgaatgt 1525
QY 1259 tgattcaaatgcttgcagagaagaagt 1283
Db 1526 cctcttcaagctttagaagaataact 1550

```

RESULT 6
AAC62134
ID AAC62134 standard: DNA: 2128 BP.

XX AAC62134;

XX 06-MAR-2001 (first entry)

DE Nucleotide sequence of beta-1-6-N-acetylglucosaminyltransferase.

```

KW Human: beta-1-6-N-acetylglucosaminyltransferase; C2GNT-M; inflammation;
KW membrane protein; branched sialyl Lex; L-selectin; immune reaction;
KW Inflammation; tissue rejection; tumour metastasis; SS.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 354..1670
FT /tag= a
FT /product= "beta-1-6-N-acetylglucosaminyltransferase"
FT polyA.signal 2100..2105
FT /tag= b
XX
XX US6136580-A.
XX
XX 24-OCT-2000.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX 19-JAN-1999; 99US-0233506.
XX
XX (BURN-) BURNHAM INST.
XX
XX Fukuda M, Yeh J;
XX
XX WPI: 2001-040238/05.
XX
XX P-PSDB; AAB30518.
XX
XX New C2GNT-M polypeptides having core 2, core 4 and I branching
XX beta-1-6-N-acetylglucosaminyltransferase activities for preparing
XX reagents useful for diagnosing, preventing or treating inflammation or
XX tumour metastasis
XX
XX Example 1; Fig 4; 25pp; English.
XX
XX The present sequence encodes a human
XX beta-1-6-N-acetylglucosaminyltransferase polypeptide, with core2, core4
XX and I branching activities. It is designated C2GNT-M. C2GNT-M is a
XX membrane protein that is predominantly expressed in colon, small
XX intestine, trachea, stomach and thyroid, as well as in certain cancer
XX cell lines. C2GNT-M polypeptides may be used to prepare molecules having
XX highly branched sialyl Lex and L-selectins, which may be subsequently
XX used to modulate immune reactions, e.g. inflammation and tissue
XX rejection, and to prevent or inhibit tumour metastasis.
XX
XX Sequence 2128 BP; 569 A; 477 C; 526 G; 556 T; 0 other;
SQ

```

Query Match 14.1%; Score 191.8; DS 22; Length 2128;
Best Local Similarity 52.5%; Pred. NO. 7.3e-44;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY 317 tgacaagtattgtgacattatcacagactcgaaggttattgtctcaaaagctgtctcaa 376
Db 670 tcacacagagcttggagcacttcaagctgaaggaagttcatacagttcccactgagca 729
QY 377 agggagggaaagagtttcccaatagcctatcttgtgtgtgcccaagaagatgaatattg 436
Db 730 aagaagaggtgaggttccctatgcatcactcattgtggtatcattgaaatgtgaact 789
QY 437 ttgaaagccttaccatgcttatatacaacccagacataatttaccatcatcatgatac 496
Db 790 ttgaaagccttaccatgcttgcgtgtatgtccctccagaaacataactgtgtccattgtgagt 849
QY 497 gtaaggaaccttgaaccttcaaaagttggccatggaacaatttgaatgcttcccaata 556
Db 850 agaagtcctcccaagaaccttcaaaagcggttcaaaagaacatttcttcttcccaaatg 909
QY 557 ttctattgttccaaattagagagcttggaaatagccacatttccaaagctcgaagctg 616
Db 910 tcttcatagccagtaaaactgtgtccggtgttattgtctcctcctgttccaggtgtcagctg 969
QY 617 atttaattctgttgcgaaccttccgaagcttcccaatccagtggaataatgttataact 676

```

```

Db 970 acccaactgcatggaagatctgtccagagctcagagtcggtggaatacttctcgata 1029
QY 677 tctgtggcgaagatttccctcgtgaagccaatttgaattggtgctcgaattgaaaaac 736
Db 1030 catgtggagcagacttcccttaagaagcattgacagatggttccagcctctaagatgt 1089
QY 737 tcaatggagcaaatgtgtgagacggtgaaaccccccaacagtaaatgtgaaagattca 796
Db 1090 tgaatggaggaatagatgagtcagtcagagtaacctcctaagacacaaagaacccgctgga 1149
QY 797 ctacccatcatgaacttaagacggtgctcctatgaatagttgaagctaccaataagacaa 836
Db 1150 aatccaacttggagtagtgagagacacattacac-----ctaacca 1191
QY 857 acatcccaagggaagcaccocccataacattcagatattgttggcagtgctattttg 916
Db 1192 acaagaagaagagatcccccctataatttaactatgtttacaggaatgctgtaattg 1251
QY 917 tttaagtcgaagcattgtttaataatatttcaacaactccatcgttcaagactttttg 976
Db 1252 tggctcccgagattcgtccaacatgttttgaagaacctaaatcccaacactgattg 1311
QY 977 cctgtcttaagacacatactctcctgatagacacttttggctcacttgatcgggttc 1036
Db 1312 aatgggtaaaagacactatagccagatgaacacctggtggccaccttcacagctgac 1371
QY 1037 caagaaactcgtggaggaattccacagacacagccagagtg--tgtcgtactgcaagata 1093
Db 1372 gttgagtcgctgctcgtctccacacaccccaagtaagcactctcagatgaactcta 1431
QY 1094 agactcgcctgtcaagtgtgaattacatgaagcttttctacccagt----- 1143
Db 1432 ttgcagagctgtgtaagtgaggtgcatagggagacatcgataaagggtgctcctatg 1491
QY 1144 -----tgcagtgatccacacttcgaagcgtgttattttgagcgcgcgaataagct 1198
Db 1492 ctccctgctcgtgataccacacagcggtatctgcttattggcgctggggaactgattc 1551
QY 1198 ggcattacaagatgagacattgttcttaataaattgattcttaagggtggaacctctc 1258
Db 1552 ggaatgctccaacacatcacctgttggccacaagattgaccccaagtagatgataatg 1611
QY 1259 tgattaatgcttgccagaaagct 1283
Db 1612 ctcttcagtgcttagaagaatact 1636

```

```

RESULT 7
AAC99109
ID AAC99109 standard; cDNA: 2229 BP.
AC AAC99109;
DT 09-MAR-2001 (first entry)
XX
XX Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:337.
DE
XX
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytosolic; neuroprotective;
KW neotrophic; immunomodulatory; relaxant; contractile; gynecological;
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
XX Homo sapiens.
OS
XX WO200055320-A1.
PN
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05989.
PF

```

```

XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-579444/54.
DR
XX P-PDB: AAB54344.
PT
XX New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particularly pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
XX Claim 1; Page 759-760; 1379pp; English.
PS
XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytosolic,
CC neuroprotective, neotrophic, immunomodulatory, relaxant, contractile,
CC gynecological, cardiant and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
XX Sequence 2229 BP; 605 A; 501 C; 546 G; 572 T; 5 other;
SQ

```

```

Query Match 14.1%; Score 191.8; DB 21; Length 2229;
Best Local Similarity 52.5%; Pred. No. 7.4e-44;
Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

```

QY 317 tgaccagatgtgtgacattatcagacrtcgaaggtatgtctcaaaagctgtctca 376
Db 743 tcaccagagactgtgagcacttcaaggtcgaaggtacatgcttccactgagca 802
QY 377 aggaagagaagaagcttcccaatagcttcttgtgtgcacaaagatgcaattatg 436
Db 803 aagaagaagtgaggttccctatgcatctcattgtgttcatagagaatgaaact 862
QY 437 ttgaaggttalcattcctatatatacaccagacataattacgtacatcatatgac 496
Db 863 ttgaaggtctactgcaggtgtgtatgctccctcaagaacatactgtccatgtgagtg 922
QY 497 gtaagcacctgtactcctcaaaagtgtccatgaacaattagctaaagtcttccata 556
Db 923 agaaagccccaagaacttccaaagagcgttcaagaacattatcttcttcccaag 982
QY 557 ttctcattgttccaaattagaggtgtggaatgtccacattccagactcaggtct 616
Db 983 tctctatagccagtagcgtgtcgggtgtgtatgtatccctccggtccaaaggtgacgtg 1042
QY 617 atttaattgctgtgcgacacttgaagcttccatcagtcagtggaataatgttatcaat 676
Db 1043 acctcaactgcatggaagacttctccagagctcgaatgctcggtgaataatcttccgaata 1102
QY 677 tctgtggcgaagatttccctcgtgaatatttgaattggtgtcagaagtgtgaagaac 736
Db 1103 catgtggagcagacttccataaagaagcaatgcagagatggtgcaggtctccaagatgt 1162

```

QY	737	tcattatgagcaaatatgtttggagacggtgtgaacccccaacacgtataatctggaaagtctta	796
Db	1163	tgatagtgaggaataatgacatggagtcagagatgacctcttaagcacaagaacccgctgga	1222
QY	797	cttaccctcatgtaacttgaagcagggtgtgcttatgatatgtatgtagagctacataagtagaca	856
Db	1223	aattcaactttgtagtgtagagagacattacac-----ctaacca	126
QY	857	acatctccaaggaagcaccoccccatactcttcagatattgtttggcagtgctattttg	916
Db	1265	acaagaagaagatccctcccccctataatttaactatgcttacaagggaatgcgtacattg	132
QY	917	ttttaagtcagaacatttgtttaatatatttccaacaactcatcgttctcaagactttttg	976
Db	1335	tgcttcccgagatttcgtccaacatgttttgaagaacccctaactccacaacatcgtatg	138
QY	977	ccctgtctaaagaacactactctctctatgtagacattttgggtctacttgatctgggttc	1033
Db	1385	aatgttgtaaaagacactataagccccagatgaacaacaccttgggcacccttcagcgtgcac	144
QY	1037	caggaaataacctggggagatcttccagatcagccccagagt--tgctcatctgcagagta	109
Db	1445	ggttgatgctgcgtgctgtctgttcccaaccccccaagtagagacatctagaacatgactcta	150
QY	1094	agactgcctgttcaatgtgaatatactatagtaaagcctttttctatccagtt-----	114
Db	1505	ttgcacagccttggtcaagtgcagaggtcatatgaggagacatcgtataagggtgctccctatg	156
QY	1144	-----tttactggaatccacaccttcgaagcggtgtgtattttatggagctgcagaattaggt	119
Db	1565	ctccctgctcgcgatactccaccacgagcgctactcgcgtttatgggtgtggagactgaatt	162
QY	1199	ggcttataccaagatggaacattgttgcctataataatttgattcttaaggtggaacctatct	125
Db	1625	ggatgcttccaacaaacacactactgttggccaacaaagttagcccaaggtatgataatg	168
QY	1259	tgattaatgtcttgcagaaagcct	1283
Db	1685	ctcttcagtgcttgaagaataact	1709
RESULT 8			
AAH34463			
ID	AAH34463 standard; cDNA; 2236 BP.		
XX			
AC	AAH34463;		
XX			
DT	03-SEP-2001 (first entry)		
XX			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:1545.		
XX			
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; Chromosome 15; ss.		
OS	Homo sapiens.		
XX			
XX	WO200122920-A2.		
XX			
PD	05-APR-2001.		
XX			
PF	28-SEP-2000; 2000MO-US26524.		
XX			
FR	29-SEP-1999; 99US-0157137.		
FR	03-NOV-1999; 99US-0163280.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
WP	1; 2001-235357/24.		
XX			
PR	P-PSDB; AAG75058.		
XX			

```

Db      1332  tggctccgcgagattctgtccacacatgcttttgagacacccctaataatcccaaacatgattg 1391
QY      977  cctggtctaaagacacatactctctcgtatgagacacttttgggctacttgattcgtgttc 1036
Db      1392  aatgggttaaaagacactatagaccagatgaacacctctgggacaccccttcagcgtgcac 1451
QY      1037  caggaatacctctgggagattccagatccagccagatg---gtctgtcttcagagta 1093
Db      1452  ggtgagatgctgtctgtctgtcccaacaccccaagtaagacatctcagacatgacttcta 1511
QY      1094  agactgcgcttgcgaagtgaattactatgaagcttttctatccagc----- 1143
Db      1512  ttgcagagctgctgaagtgcaggtatagggagacacgtataggggtgtcctctatg 1571
QY      1144  ----tctactggtatccacacttcgaagcgtgtgtattatgagcttcgaagtaagt 1198
Db      1572  ctccgtgctctgtaattccaccacagcggcctatctgccttctatgggctggggaactta 1631
QY      1199  ggcttatcaagaagtgcacattgtgttctaataattgattcttaaggtggacctatct 1258
Db      1632  ggaatgcttcaaacaccatcacctgttgccacaagtttgacccaagtgatgataatg 1691
QY      1259  tgaattaatgcttgcgcgaagaaagct 1283
Db      1692  ctcttcagtgcttagaagaataacct 1716

RESULT  9
ID      AAA48623
AC      AAA48623:
DT      19-SEP-2000 (first entry)
DE      Human C2/4Gnt CDNA.
XX      Human: C2/4Gnt; UDP-N-acetylglucosamine; O-glycan biosynthesis;
KM      O-glycan beta-1,6-N-acetylglucosaminyltransferase; cancer; ss.
XX      Homo sapiens.
XX      Key      Location/Qualifiers
FH      CDS      496..1812
FT      /*tag= a
FT      /product= "C2/4Gnt"
FT      primer_bind 586..605
FT      /*tag= b
FT      primer_bind 1794..1813
FT      /*tag= c
FT      polyA_signal 2244..2249
FT      /*tag= d
XX      WO200034449-A2.
XX      15-JUN-2000.
XX      03-DEC-1999; 99WO-DK0677.
XX      04-DEC-1998; 98DK-0001605.
XX      (CLAU/) CLAUSEN H.
XX      Clausen H, Schwientek T;
XX      WPI; 2000-423407/36.
XX      P-PSDB; AAY94492.
XX      New nucleic acid molecule encoding UDP-N-acetylglucosamine useful as
PT      probe for the detection of specified glucosaminyltransferase from
PT      other species and related organisms
XX

```

```

PS      Claim 5; Fig 2; 47pp; English.
CC      The present sequence encodes human UDP-N-acetylglucosamine:
CC      N-acetylglucosamine beta-1,6-N-acetylglucosaminyltransferase
CC      (C2/4Gnt). The protein is the third member of the family of O-glycan
CC      beta-1,6-N-acetylglucosaminyltransferases. It adds core 2 or core 4
CC      based O-glycans on to oligosaccharides, glycoproteins and
CC      glycosphingolipids. C2/4Gnt can therefore be used in the production of
CC      appropriately glycosylated glycoconjugates with particular enzymatic,
CC      immunogenic, or other biological or physical properties. The nucleotide
CC      sequence is useful as a probe for the detection of C2/4Gnt from other
CC      species and related organisms and for the recombinant production of
CC      C2/4Gnt polypeptide. The nucleotide sequence was identified by analysis
CC      of EST database sequence information. Oligonucleotides derived from EST
CC      clone 17865 of ATCC were used to isolate two full-length C2/4Gnt clones
CC      from a human foreskin genomic pl library by 5' RACE PCR. RT-PCR was
CC      performed using Colo205 human cell line mRNA in order to produce cDNA
CC      for expression of C2/4Gnt in Sf9. The control of O-glycan core assembly
CC      has been implicated in tumour progression and metastasis.
SQ      Sequence 2319 BP; 615 A; 528 C; 559 G; 617 T; 0 other:

```

Query Match 14.1%; Score 191.8; DB 21; Length 2319;

Best Local Similarity 52.5%; Pred. No. 7.5e-44; Matches 517; Conservative 0; Mismatches 432; Indels 36; Gaps 3;

```

QY      317  tgaccagatgattgacattatcatcagacttaagaaggtatgctcaaaagctgtctca 376
Db      812  tcaccagagactgtgagcacttcaagctgaaggaagttcatcagcttccacatgagca 871
QY      377  aggaagagaagaagcttcccaatagcctattcttgggtgtccacaagaatgcaattatg 436
Db      872  aagaagaggtggaagtcctcatgcatcctcatcattggtatcagatcagagaagattgaaact 931
QY      437  ttgaaggttcatcattcatatatacaacccagacacataattcatgcttccattatgctc 496
Db      932  ttgaaggttactgtagctgtgtatgctccctcagacaatactatgcttcatgtgagtg 991
QY      497  gtaaggaacctatattccttcaaatgttgcacatgaacaaattgtcgaagtgcttcccaata 556
Db      992  agaagtcctccagaacttccaaagcgtgtcaagaagcaattattctgtctcccaatg 1051
QY      557  ttctattgcttcccaatagaagcgtgtggaatagcccaattccagatccagatcgtc 616
Db      1052  tcttcattagccagtagctgtgtcggtgttattagcttccctcgtgcaggggtcgaagctg 1111
QY      617  atttaattgctgtcgcgaacctctgaagcttcaatcccaagtgaacatattatcaact 676
Db      1112  acctcaactgcatggaagactgtctccagagctcagtgccgtggaataacttctcgtatba 1171
QY      677  tgtgtggcgaagatttcccccgaagtcgaatttgaattgtgtgcagagttgaaaaac 736
Db      1172  catgtgggacgacttcttcataaagagcaatgacagatggtccagctctcaagaagt 1231
QY      737  tcaatggagcaaatatgttggagcgtggaacccccaacacagtaaatgtgaaagatca 796
Db      1232  tgaatggagagatagcatgagtgatcagagtgctcccttaagcacaaaccccgctgga 1291
QY      797  cttaacatcatgaaacttagcagcgtgtccttatgaaatgtgaagtgaacataagagca 856
Db      1292  aatatcaactttaggtatgtggaagacattcac-----ctaaacca 1333
QY      857  acatctcaagaagaagaccccccataacatcagaatatgttggcagtgcttatttg 916
Db      1334  acaagaagaagatctcctcccttataatttaactgtttacagaggaatcgttactatg 1393
QY      917  tttaagtaagcattgtttaaataattatccaacacccatcgttcaagactttttg 976
Db      1394  tggcttccggaatttgcaccaacatgttttgaagaaccctaaatcccaacactgattg 1453
QY      977  cctgttctaagaacacatactctctgtatgagcaatttgggctacacttgattcgggttc 1036

```

Db 1454 aatggtataaagacattatagccagatgacacccctctggccaccccttcagcgtgcac 1513
 Qy 1037 cagaataactcgtggagatctccagatcagcccgagatg---gtctgacttcagagta 1093
 Db 1514 ggtggagcctgctcgtctgtcccaacaccccaagtgacacatctcagaatgactcta 1573
 Qy 1094 agactcgccttgcataagtgagattactatgaaggcttccttcacagc-----1143
 Db 1574 ttgcagagcgtcgtcaagtgagcgggtcattggagagacatcagataaagggtccctatg 1633
 Qy 1144 -----tgactcgccttcaccccttcgaagcgtgtattttgagcgtgcgaattaaagt 1198
 Db 1634 ctccctcgtcctgcagatccacccagcggctatctcgtttatggggcgtggacttgatt 1693
 Qy 1199 gactatcaagaatgagacatcgttctgataatgataattgataagtgagccctatct 1258
 Db 1694 ggaagcttcaaaacacatcactcgttggcacaagaattgacccaagaagtatgataatg 1753
 Qy 1259 tgattaatgcttgcgcagaaagct 1283
 Db 1754 ctctcagtgcttagaagaatacct 1778

RESULT 10
 AA089201
 ID AA089201 standard; cDNA; 1807 BP.
 AC AA089201;
 XX
 DT 23-SEP-1995 (first entry)
 XX
 DE I-branching enzyme cDNA.
 XX
 KW Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGF;
 KM inflammation; antiinflammatory; tumor; hypersensitivity; anemia;
 XX transgenic animal: I antigen; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 255..1457
 FT CDS /*tag= a
 XX
 PN M09507020-A.
 XX
 PD 16-MAR-1995.
 XX
 PF 09-SEP-1993; 93WO-US08476.
 XX
 PR 09-SEP-1993; 93WO-US08476.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 XX
 PI Bierhuizen MFA, Fukuda M;
 XX WPI; 1995-123182/16.
 DR P-PSDB; AAR71932.
 XX
 PT Nucleic acid encoding I-branching enzyme - used to develop prods.
 PT for treating e.g. inflammatory responses, tumors or
 PT hypersensitivity reactions
 XX
 PS Claim 1; Fig. 6; 56pp; English.
 XX
 CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was
 CC reverse-transcribed and inserted into expression vector pCDNA1.
 CC Plasmid DNA was used to transfect CHO-Py-lau cells. Transfected
 CC cells were screened using human anti-I antigen antibodies and goat
 CC anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in
 CC AA089201) encoding I-branching enzyme (AAR71932).
 XX
 SQ Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;

Query Match 13.6%; Score 185.2; DB 16; Length 1807;
 Best Local Similarity 52.1%; Pred. No. 4.9e-42;
 Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

Qy 327 ttgtgacattatcagacctaagaaggtatgctcaaaaagctgtctcaagaagagagaa 386
 Db 467 ttgcaagagatctctgaccacagagccacactacatcaacagcccttattctcaagaagagc 526
 Qy 387 aagcttcccaatagcctattcttgggtgtgcacaaagatgcgaattatggttaagagct 446
 Db 527 tgacttcccttggacatataatgttcatccatcatcacttgcacacttgcagaagct 586
 Qy 447 tatccagctatataacacagcacacataattaccgtcatcattatgatacgtgaagcacc 506
 Db 587 ctccagggctatttaaatgtcccaaatctcactcgtgtctcgttgatgataaagcaac 646
 Qy 507 tgataccttcaaaagttgcagataacaatttagctaagtgcttcccaattttcattgc 566
 Db 647 aactgaatttaagatgctgtagagcaacttaagctgtctcccaaacgcttctctgc 706
 Qy 567 ttccaatatagagcgtgttgaaatgacccacatcttcacagctccagcgtgattaatg 626
 Db 707 ttccaaagatggaacccgttgcatacagaggtcccaagctccagcgtgacctgaactg 766
 Qy 627 ctgttcgaccttctgaagcttccaatccacagtggaatagttatacacttgttgggca 686
 Db 767 catcagagatcttctgccttcaggtctcagatggaagtagtatacacaaccttgggca 826
 Qy 687 agatttccctcgtgaatcaaatttgaattgtgtcagagtttgaanaaactcaatgagc 746
 Db 827 agacttccctcgtgaatcaacaacaagaataagtaattcgaagaagatttaaggtaa 886
 Qy 747 aaatatgttggaagcgttzaaaccccaacagtaattgaaagatccaattacatca 806
 Db 887 aactatcacccacaggggtgtcgtcccaagctcatgcaattgagcagactaataatgtcca 946
 Qy 807 tgaacttaagcgggtgcttataatgataatgtaagctacaataaagacacaactccaa 866
 Db 947 ccaagagcacccttggca-----aagagcttccctatgtaataaacaacagcgttaa 1000
 Qy 867 ggaagcaccocccataacatccaattgttggcagtgcttattgttttaagca 926
 Db 1001 accgcctcccccataactcaaatcttaacttgcctcgtcgtcgtatggtcctacaag 1060
 Qy 927 agcatttgcataatatttcaacaactccactcgttcaagaacttttgcctgtctaa 986
 Db 1061 agagttgcacaacttgcctcgtatgacccacagcgtgttgaattgctcagtgctcaa 1120
 Qy 987 agacacatactcctcgtatgacacacttgggcttaacctgattcgggtccaggaatacc 1046
 Db 1121 ggaacattccagtcctcgtatgacacacttggggtgacacccaataagattcccggttcc 1180
 Qy 1047 tggggagatttccagatcagcccaagatgtgtcgtatctgcagagtaagaatcgctgt 1106
 Db 1181 tggctatagccaatagcatcctcgtgactg-----gaaacctcaagcgtat 1225
 Qy 1107 caagtgaattactatgaaagcgttcttataccagtgtaactgatactccacttcgaag 1166
 Db 1226 aaggtgagatgacatggaagacagagagc---tgccagggcaataatgtacatg 1282
 Qy 1167 cgtgtattatgagcgtcgcgaattaaagtggtctatcaagaatggaacattgtttgc 1226
 Db 1283 tattgtatcatgaaagagagacttaagcgtgttaattcaaccaagcctgtttgc 1342
 Qy 1227 taataatttgatctcaagtgagccctatcttgataattccttggcgaagaaacttga 1286
 Db 1343 taacagtttgagcttaataactaccccttactgtggaatgctcagaactaaggcatcg 1402
 Qy 1287 agaa 1290
 Db 1403 cgaa 1406


```

RESULT 11
AA16201
ID AA16201 standard; cDNA; 1807 BP.
XX
AC AA16201;
XX
DT 22-APR-1996 (first entry)
XX
DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme cDNA.
XX
KM Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGNT;
KW glycosyltransferase; blood group; I antigen; polylactosaminoglycan;
XX ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 255..1457
FT /tag= a
XX
PN US5484590-A.
XX
PD 16-JAN-1996.
XX
PF 09-SEP-1993; 930S-0118906.
XX
PR 09-SEP-1993; 930S-0118906.
XX
PI (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI Bierhuizen MFA, Fukuda M;
XX
DR WPI: 1996-087019/09.
XX
DR P-PSDB: AAR92474.
XX
XX
XX Human beta-1,6-N-acetylglucosaminyl transferase, I-branching enzyme
PT - used to develop prods. for the study, detection and treatment of
PT pathological conditions involving the enzyme.
XX
PS Example 4; Column 29-32; 29pp; English.
XX
CC A cDNA insert (AA16201) in pCDNA1-IGNT codes for human beta-1,6-N-
CC acetylglucosaminyltransferase, I-branching enzyme (IGNT) (AAR92474),
CC an enzyme that converts blood group I antigen to I antigen. The
CC cDNA was isolated from a cDNA library pred. from human PA-1
CC teratocarcinoma cells in vector pCDNA1 following transfection into
CC CHO-Py-leu cells and screening with human anti-I antibodies. The
CC cDNA can be used to regulate the expression of human IGNT or to
CC modify its biological function, to produce soluble or membrane-bound
CC forms of IGNT in host cells, to breed transgenic animals, and to
CC design antisense oligonucleotides.
XX
SQ Sequence 1807 BP; 511 A; 401 C; 398 G; 497 T; 0 other;

```

```

Query Match 13.6%; Score 185.2; DB 17; Length 1807;
Best Local Similarity 52.1%; Pred. No. 4,9e+42;
Matches 502; Conservative 0; Mismatches 438; Indels 24; Gaps 3;

```

```

QY 327 ttgtgacattatcagacttaagaggtatgctcaaaagctgtctcaagagagaa 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 467 ttgcaagaaatacttaccacagacacacacacacacacacacacacacacac 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 aagcttccaaatagctattcttggtgtgccaacaagtgcaatattgttgaaagct 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 527 tgccttccttgccatataatagtcacacacacacacacacacacacacacacac 586
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 tatccatgcatataacacacacacacacacacacacacacacacacacacacac 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 587 cttcagagctattacatgccccaataatctctctgttcatgtgtgataaagcaac 646
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 507 tgatacttcaaaagtgtgcacatgaacaattagcttcttcccaattttctatgc 566

```

```

Db 647 aactgaatttaaagatcggtagagcaacttaagctgtctcccaaacgctttctgcgc 706
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 567 ttccaaattagaagctgtggaatatgcccacattccagactccagcgtcaattaaatg 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 707 ttccaagatgaacccgtgtgtctatggaaggatctccaggtccagcgtcagcgaactg 766
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 627 ctgtcggaacctctgaagcttcaatccagtggaataatgtatcaactgtgtgggca 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 767 catcagaatcttctgcctcagagctcactgaagtaagcttatcaacaccttgggca 826
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 687 agatttcccttgagaagcaaatgttgatgtgtgtcaggttggaataaacctcaatggagc 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 827 agacttcccttgagaagcaaatgttgatgtgtgtcaggttggaataaacctcaatggagc 886
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 747 aaatatgttgagaagcgttgaaacccccaacagtaaatgtggaagattcacttaccatca 806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 887 aaatatcccccaaggggtgctgcccacagctcacttgagcgtcaataatgttcca 946
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 807 tgaacttgaagcgtgtgcttctatgaaatgttgaaatcccaataagcaaacatctccaa 866
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 947 ccaagagacacttgagca-----aagagcttccatgtatgataaagaacagcgttgaa 1000
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 867 ggaagcaccgccccaataacattcagataattgttgcagtgctatttgtttaaagca 926
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 accgcctccccccataatctcacaatttacttggcttgcctatgtgtgtgtatcaag 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 927 agcatttgttaatatatttcaacaactccatcgttcaagacttttgcctgtgtctaa 986
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 agagtttgcacacttgttctgcacatgaccccggtgtgtgttcttccagtggtccaa 1120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 987 agaacatattctcctgatgagacacttgggtgacacttgatccgggttccaggaatacc 1046
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 ggacacttcaagctcctgatgagacacttctcgtgtgacacactcaatgagtgatccaggtgtcc 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1047 tgggagagcttccacagacccacagagatgtgtcgtactcgaagagtaagactgcctgt 1106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 tggctctgtccaaatgcatctcgtgactg-----gaaacctcagaagat 1225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1107 caagtgaattactacatgaagcgttctctatccaggtgtgactgacatccactcgaag 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1226 aaagtggagtgatcagtgagacacagacagcagc---tgcaacgycacactgtgaactcgtg 1282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1167 cgtgtgttattatgagctgcagaataaaggtgtgcttatcaagaatgacatgtgttc 1226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1283 tattgttcatgtgaaacgagacacttaaggtgtgtgttaattcaccaagcctgtttgc 1342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1227 taataaattatcttaaggtgagccatcttgaatgaatgaatgtgtgcaagaagcttga 1286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1343 taacaagtttgagtttaattactacccttaccgttgaaatgactgaacacagagcatcg 1402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1287 agaa 1290
   |||
Db 1403 cgaa 1406

```

```

RESULT 12
AAV30006
ID AAV30006 standard; cDNA; 1807 BP.
XX
AC AAV30006;
XX
DT 11-AUG-1998 (first entry)
XX
DE Full length cDNA sequence of human IGNT.
XX
KM Human: beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme;
KW IGNT; core 2-beta-1,1-N-acetylglucosaminyltransferase; C2GNT;
KW treatment; disorder; under expression; type-II hypersensitivity reaction;
KW neonatal haemolytic disease; autoimmune haemolytic anaemia;
KW thrombocytopenia; ds.
XX
OS Homo sapiens.

```


Fri May 3 10:57:52 2002

us-09-645-192-1.rng

Page 15

Db 1211 caagagtgctccctaccgcgcctcgatvgagtcattgtgctccagtgtgcatlccg 1370

Oy 1182 agctgcgaattaaagtgtgctctacaagaatgacattgtttctcataaattgtatc 1241

Db 1371 agcgcgtgacttgaactacggatgctgcgcaaacacacttggtttgcataaagtgtgact 1430

Oy 1242 taaggtggaacctactctgtatcaaatgctctgcaga 1277

Db 1431 ggaagtggaccttcttgcacacgaagtgttgatata 1466

Search completed: May 2, 2002, 11:37:34
Job time: 7474 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 08:18:30 ; Search time 1511.47 Seconds
(without alignments)
9683.123 Million cell updates/sec

Title: US-09-645-192-1
Perfect score: 1362
Sequence: 1 atgaagatttcacatgtta.....atctaccatccacatcarga 1362

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estom: *
5: em_estpl: *
6: em_estro: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	437.8	32.1	821	11 BG434801	BG434801 602507322
2	408.4	30.0	410	13 AQ005888	AQ005888 CIT-HSP-2
3	393.2	28.9	612	13 AQ000348	AQ000348 HS_2175_B
4	350.8	25.8	590	13 AZ603057	AZ603057 IM0122E09
5	289	21.2	292	13 AQ091453	AQ091453 HS_3016_B
6	253	18.6	908	13 CNS02A46	CNS02A46 Tetraodon
7	221	16.2	973	13 CNS058A	AL188089 Tetraodon
8	187	13.7	1367	13 AK019924	AK019924 Mus muscu
9	174.2	12.8	1862	12 AK008234	AK008234 Mus muscu
10	170.8	12.5	860	12 AL576150	AL576150 AL576150
11	163.2	12.0	970	13 CNS04UIM	AL107831 Tetraodon
12	157.6	11.6	756	13 CNS01UWZ	AL168380 Tetraodon

C 13	148.8	10.9	526	10 BE005007	BE005007 MRO-BN011
C 14	148.8	10.9	612	10 BE292814	BE292814 601105179
C 15	142	10.4	873	11 BG185546	BG185546 RST4485_A
C 16	139.4	10.2	567	13 AZ388491	AZ388491 TM0148G03
C 17	136.6	10.0	798	11 BG220853	BG220853 RST40650
C 18	135	9.9	774	11 BG204638	BG204638 RST24051
C 19	133	9.8	572	11 BE234997	BE234997 142220_MA
C 20	133	9.8	655	10 AM842622	AM842622 MR2-CN003
C 21	129.8	9.4	1050	11 A1722764	A1722764 fc31c11.Y
C 22	127.8	9.2	869	10 BE867668	BE867668 601443206
C 23	124.8	9.2	869	10 AL555400	AL555400 AL555400
C 24	121.6	8.9	820	11 BG198987	BG198987 RST18281
C 25	118.2	8.7	695	11 BF036111	BF036111 601457742
C 26	116.2	8.5	842	11 BE869192	BE869192 601445191
C 27	115.8	8.5	554	10 AA307800	AA307800 EST178656
C 28	113.8	8.4	671	10 AM233232	AM233232 f129g12.Y
C 29	110.8	8.1	788	10 A1528293	A1528293 u195f11.Y
C 30	110.8	8.0	791	11 BG788203	BG788203 SEMC008
C 31	108.8	7.9	428	13 AQ437509	AQ437509 HS_513_B
C 32	102.4	7.5	506	11 B1337733	B1337733 361252_MA
C 33	102.4	7.5	348	11 BE863271	BE863271 UI-M-BHO-
C 34	101.2	7.4	482	11 BF416888	BF416888 UI-R-CNO-
C 35	99.6	7.3	542	10 AW388627	AW388627 MR2-ST013
C 36	97	7.1	633	10 A1133042	A1133042 HA1630_Hu
C 37	96.6	7.1	398	13 A2149281	A2149281 SP_0035_B
C 38	95.8	7.0	622	10 A1747955	A1747955 u104d07.Y
C 39	93.4	6.9	605	13 A1194803	A1194803 SP_1028_B
C 40	92.4	6.8	555	10 BE721246	BE721246 188355_MA
C 41	90.6	6.7	443	11 BE755013	BE755013 QVO-CT058
C 42	88.6	6.5	489	10 BE721377	BE721377 188533_MA
C 43	88.2	6.5	438	10 AM842590	AM842590 MR2-CN003
C 44	87	6.4	543	11 BR077138	BR077138 225944_MA
C 45					

ALIGNMENTS

RESULT 1
LOCUS BG434801
DEFINITION 602507322F1 NIH_MGC_79 Homo sapiens CDNA clone IMAGE:4604659 5',
mRNA sequence.
ACCESSION BG434801.1 GI:13341307
VERSION BG434801
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
1 (bases 1 to 821)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLOMTECH Laboratories, Inc.
CDNA Library Preparation: CLOMTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMU at:
http://image.llnl.gov
Plate: LCM1340 row: k column: 06
High quality sequence scoop: 709.
Location/Qualifiers
1..821
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4604659"
/clone_lib="NIH_MGC_79"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Placenta; Vector: pNNR-LIB (Clontech);

FEATURES

source

REFERENCE
AUTHORS
1 (bases 1 to 612)
Mahairs,G.G., Wallace,J.C., Smith,K., Swartzel,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. a

TITLE
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE
99380589
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2175 Row: B Column: 23
Seq primer: T7
Class: BAC ends
High quality sequence stop: 612.
Location/Qualifiers

FEATURES
source
1. 612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIR Approved Human Genomic Sperm Library D"
/sex="Male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"

BASE COUNT 180 a 141 c 121 g 165 t 5 others
ORIGIN

Query Match 28.3%; Score 393.2; DB 13; Length 612;
Best Local Similarity 91.6%; Pred. No. 9e-91;
Matches 427; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 839 agctcaataaagacacacatctccaaagagaccccccaataacattcagatttg 898
Db 610 AGCTCCAAATGAGAGTCAAGCTCTCCAGAGACACCGCTCATACATTCAATATG 551
QY 899 ttggc-agtgcattattgttttaagcagcattgttaataataatttcacaactcc 957
Db 550 TTGGCAGTGTATTTTGTTCAGTCAGCATGTGTTAAATATTTCAAACTATC 491
QY 958 atggttaagactttttgtccttgctcaagaacacatactcctcgtatgagcatttg 1017
Db 490 ATGCTTAAACCTTTGTGGTGTGAGACACATCTCATATGAGCACTTGTG 431
QY 1018 gctacattgattcggttcaggaataacctgaggagatttcagatcagccagagtgc 1077
Db 430 GTTACCTTGATTCGGGTTCCAGAAATACCTGGGAGATTCCAGATCAGCCAGATG 371
QY 1078 tctgattcgaagtaagactgcctgttcaagtgaattactataaaggctttctat 1137
Db 370 TCTGATCTGAGAGATGAGATCCCTGTGTCAGTGAATTAATTAAGCTTTTCTAT 311
QY 1138 cccagttgactgattcacccttcgaagagtggtfattaatgagctcagaattaag 1197
Db 310 CCCAGTTGATCGATCTCACCTTCGAGAGCTGTGTAATTAAGAGCTGACGATTAAG 251
QY 1198 tggctatcaagaatgacattggtttgctaataaatttgattcctaagtgagccatc 1257
Db 250 TGGCTTATCAAAAGATGACATGTTGTTGTAATAATTGATCTAAGGAGACATATC 191
QY 1258 ttgattaaagcttgcgagaagaagcttgtaagaacagagaagact 1303
Db 190 TTGATTAATATGCTTGGAGAAAGCTTGCGATGCGTCCAGGTGCGCT 145

RESULT 4
A2603057 590 bp DNA GSS 13-DEC-2000
LOCUS AZ603057
DEFINITION 1M0422E09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0422E09 F, DNA sequence.

ACCESSION
A2603057
VERSION
GI:11725247
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 590)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tinger, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
64112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0422 row: E column: 09
Seq primer: CGTCTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 590.
Location/Qualifiers

FEATURES
source
1. 590
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0422E09"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g114732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance.

BASE COUNT 157 a 147 c 138 g 147 t 1 others
ORIGIN

Query Match 25.8%; Score 350.8; DB 13; Length 590;
Best Local Similarity 79.6%; Pred. No. 7.6e-80;
Matches 440; Conservative 0; Mismatches 108; Indels 5; Gaps 2;

QY 1 atgaagatatacaaatgtattttaaataaccctacagagaagtttcatcctgtt 60
Db 38 ATGAAGATATTCAGATGTGCTTTAAATACACTGCCAGAGAACTTCTATCTCTC 97
QY 61 ttaaccctatgctctctcttcttgaagcttctaattg---agaagcattcttcg 117
Db 98 TTAACCTGTGCGCTTCTCTTGTGAAGCTCTTAATATGAGGAGGCTCTTCTCCT 157
QY 118 caaaagacattacttgtagtactcctaagtaacctgccttcttgtaagaacaga 177

Db 158 CAAGAGACATTACTAGTGTGATGATCCCTAAGTACATCACCATTGTGAGACAG 217
 QY 178 taaccacatgtaaggataaagcagatgtaagtaactgttcggtatcatgaacg 237
 Db 218 TTCCCGAGCTGCGGATGACAGCAGGACACAGTTAATGCTCGGGGTCTACAGCAC 277
 QY 238 gaagccttggaaatggaaagctctgaaataaagaagaagacatcaacttgag 297
 Db 278 GAGCCTTTGGAATCGCAAGAGCTAGAAATCAGAAAGCGAGCATCATCTGGAG 337
 QY 298 gatgagatgttggcaagacagacatgttgacattatgaacttaagaagttat 357
 Db 338 GAGGAGATGCTGGGATGACAGAGTGTGAGCTGATGAGCTTATGAGCCTTAAGCGATG 397
 QY 358 gctcaaaagctgtctcaaaagaggaagaaagcttcccaatagcctatcttggttgc 417
 Db 398 CATGAAAAGCTGGTTTCAAGAGGAGAAAGAGACTCCCATATGCTGTGCTGTC 457
 QY 418 cacaagaatgcaattatggttga--agccttccatgtatataaccagcacaata 475
 Db 458 CACAAAGATCCATTAGTGTGAGCGGCTTGAATCCGAGCTATNTCAACACAGACAC 517
 QY 476 ttactgacccattatgacgtcaaggcactatataccttcaagttgccaataa 535
 Db 518 TTTCCTGATCCATTATGACTGATGACACCGGACAGCTTCAAGCTGCCATGACAC 577
 QY 536 taagtaagtgctt 548
 Db 578 TAGCTAAGTGCTT 590

RESULT 5
 LOCUS A0091453/c 292 bp DNA GSS 26-APR-1998
 DEFINITION HS_3016_B1.A10.T7 CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3016 Col=19 Row=B, DNA sequence.
 ACCESSION A0091453
 VERSION A0091453.1 GI:3460364
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 292)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3016 row: B column: 19
 Class: BAC ends
 High quality sequence stop: 292.
 Location/Qualifiers
 1..292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="3016 Col=19 Row=B"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC1; BAC Clones in
 F-Coli DH10B"
 BASE COUNT 101 a 62 c 57 g 70 t 2 others

ORIGIN

Query Match 21.2%; Score 289; DB 13; Length 292;
 Best Local Similarity 99.3%; Pred. No. 5,8e-64;
 Matches 289; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 892 atattgttgcagtgcttattgttttaagtaaacattgttaataatttcaac 951
 Db 291 ATATTGTGGCAGGCTTATTGTGTTAAGTCAAGCATTTGTAATATTTTCAAC 232
 QY 952 aactcacgttcaagactttttgctgtgtcctaagaacatactcctcgtatgaac 1011
 Db 231 AACTCAGCTTCAAGATTTTGGCTGCTGCTTAAGACACATATCTCCGATGAGACAC 172
 QY 1012 ttttggctacccttgatcgggttcccaagaatccttgggagatttccagtaagccag 1071
 Db 171 TTTTGGGCTACCTTATTCGGGTTCCAGGAATACCTGGGAGATTTCAGATCAGCNCAG 112
 QY 1072 gatgtctgacatcagagtaagactcgctgtcgaagtgaatlaactatgaagcctt 1131
 Db 111 GATGTCTGATCTCAGACAGTAAAGACTCGCCTGTGCAAGTGAATTAATGAAGCCTT 52
 QY 1132 ttctatccagttgactgacatcctcgaagcgtgtattttatga 1182
 Db 51 TTCTATCCAGTTTACTGATGATCCTTCAGACCGTGTATTTATGA 1

RESULT 6
 LOCUS CNS02A4G/c 908 bp DNA GSS 12-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 251C11 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL188089.1 GI:7826193
 VERSION AL188089.1
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 908)
 Roest-Crollius,H., Tallon,O., Dasilva,C., Fizes,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 908)
 Roest-Crollius,H., Tallon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizes,C., Wincker,P., Brotier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 908)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..908
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="251C11"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG251AB06SP1-end ;
 PUC-ori"

FEATURES

source
 1..908
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="251C11"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG251AB06SP1-end ;
 PUC-ori"

BASE COUNT	156 a	266 c	283 g	188 t	15 others
ORIGIN					

KEYWORDS GSS; genome survey sequence.
SOURCE *Tetraodon nigroviridis*.

300KLE
300KLE

ORGANISM

[illegible]

REFERENCE
AUTHORS
1 (bases 1 to 973)
Roest-Crollius, H., Tallon, O., Dasilva, C., Fizames, C., Fisher, C.

TOPICAT	TITLE
freshwater pufferfish tetraodon nigroviridis	Characterization and repeat analysis of the compact genome of the freshwater pufferfish tetraodon nigroviridis

REFERENCE 2 (bases 1 to 973)

Bernot, A.; Fizames, C.; Wincker, P.; Brottier, P.; Quettier, F.; Noé, C.; Collin, H.; Dasilva, C.; Boudreau, L.; Fisher, C.;

ENTRY

1
2
3
4
5

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 073)

AUTHORS Genoscope.

JOURNAL COMMENT
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a larger

genome. For more information, please visit <http://www.genoscope.cns.fr/Tetraodon>

SOU
L. H. T. O. H. H.

Source

1.973

```
/db.xref="taxon:99883"
```

/ab_xrei="taxor

```
clone="039H15"
```

```
clone="039HL5"
```

```

/clone_id= R
/note="Genoscope sequence ID : C0AA039CD08A1-end : T31

```

BASE C
OPTIC N

Query Match	16.28; Score 221; DB 13; Length 973;
-------------	--------------------------------------

Best Local Similarity 59.78; Pred. No. 2,3e-46;

Matches 3/1; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

Qy 181 actcatgttaaggatgaaotcaagtatatgaaotiaactatcgagatatctatgaacagaa 240

25
300
200
100
0

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

241 ccltggaaatctggaagaagtcgtgaaatacgaaggacatcatctgacttgagagt 300

Db 568 CCGATGGAGGTGGGAGTCCCTTGATCATCCGGAGGAACCTCTGCCGCCGGAGTCCGAC 509

QY 301 qatqatqttqtcqaatqaccatqatttqtaacattatcaqacictaaqaattatqct 360

508

501 caaagcctgcccagaagctcccatctctgtgtccac 420

Db 448 AAGAGGTCGCTCTCGGAATGGAGCAGGACTTCCCTGGCGTACTCAATGGTTGTGCAC 389

QY 421 aaagatgcaattatggttgaaaagccttaccatgctatatacaaccaacacataatttac 480

[illegible]

201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

DB 328 TGCCTCCACTACGACCTGAAGTCCCCCTACCA GTTCATCTCGGCATAGAGGGCTGGCC 269

Db	208	AGCGGGTGAAGAACCCGACCTCAACTGCCTGTCCGAACTTCCTGAGCAGACGTAACATGG	149
Oy	661	aaataatgttatcaactctgcttggcgcaagatttcctcccttgtaagcgaatttgatggtg	720
Db	148	AAGTACCTCATCAACCTCTCGCGGCCAAGATTTCCTCCCTCAGGTCCAACAATGAGCTGTG	89
Oy	721	tccaggatgtgaaaaaacaccacatgatgagcaaatatglttgagacagtgtgaaccccccaaacagt	780
Db	88	TCAAGCTGAGAANAACATAAGCGGGCTAATATGTTAGAGAGCGCGCACCCACAGAGTAT	29
Oy	761	aaatggaagaagattcacttac	801
Db	28	AAGAGCGAGAGATTCACTTTC	8
RESULT	8		
LOCUS	AKO19924	1367 bp	mRNA
DEFINITION	AKO19924	Mus musculus adult male pituitary gland cDNA, RIKEN full-length enriched library, clone:5330430K10, full insert sequence.	HTC 05-JUL-2001
ACCESSION	AKO19924		
VERSION	AKO19924.1	GI:12860326	
KEYWORDS	CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male pituitary gland cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library clone:5330430K10.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.		
AUTHORS	1 (bases 1 to 1367)		
JOURNAL	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
PUBMED	High-efficiency full-length cDNA cloning Methods in enzymology. 303, 19-44 (1999)		
REFERENCE	99279253		
AUTHORS	10349636		
JOURNAL	2 (bases 1 to 1367)		
PUBMED	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
REFERENCE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)		
AUTHORS	20499374		
JOURNAL	11042159		
PUBMED	3 (bases 1 to 1367)		
REFERENCE	Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Komoto,H., Akiyama,J., Nishi,K., Katsuna,T., Tashiro,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,N., Nishibe,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Obara,E., Watabiki,M., Yonekura,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuiura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kirie,A. and Hayashizaki,Y.		
AUTHORS	Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		
PUBMED	20530913		
REFERENCE	11076861		
AUTHORS	4 (bases 1 to 1367)		
JOURNAL	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
PUBMED	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)		
REFERENCE	5 (bases 1 to 1367)		
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arihara,T., Carninci,P., Fukuda,S., Fukunishi,T., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hitomoto,K., Hiraoaka,T., Hori,F., Imocanti,K., Ichii,Y., Itoh,M., Izawa,M., Katoh,H., Kawai,J., Koijima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Nunakazi,R., Ohno,M., Okazaki,Y., Okita,T., Owa,C., Saito,H., Saito,R., Sekai,C., Sekai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,Y., Tejima,Y., Toyama,T., Yamamura,Y., Yasunishi,A.,		

TITLE
JOURNAL
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan. (E-mail: genome-rsg@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel.: 81-45-503-9222,
 Fax: 81-45-503-9216)
COMMENT
 Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot = 3.0 and subtraction to
 Rot = 100.0. Second strand cDNA was prepared with the primer
 adapter of sequence [5'
 GAGGAGAGATTCCTCAGTTATTAATTAATGACCCGCC 3']. cDNA was cleaved
 with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
 bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3'
 end: BamHI. Host: DH10b.
FEATURES
source
 1..1367
 location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:1912211"
 /db_xref="MGI:1925531"
 /clone="535045OK10"
 /sex="male"
 /tissue.type="pituitary gland"
 /clone.lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 235..1305
 /note="putative"
 /codon_start=1
 /protein_id="BAB31918.1"
 /db_xref="gi:12860327"
 /translation="MSLRGKFAVSALSIVIFVVFYHSQLSLPIYQGLNSSESTST
 TICDGLNHTFFETGDSPLHRLSPQRIOSHITYTSPSEEAFLPAYIWI
 KDFEERFLRAIYWPQWYCVHVDSKATDFEKVROLSCFPNAFLSKAEQVYVIT
 GESRLQDLNCKMDLVASKYVLTNGQDFPLKTRKEILNHLKRGKSKITPGVIV
 PPAYIVATVYKYVHOERKDKDGEFMKTIKTPPHQILITFGTAVVALTRFVNEII
 NDERIALLEWSKDYSPDEHFWTILNIPQVSGMPNAMSNTGLRAVKMDMEAKI
 GCHSKYVRSQNTLPVADPROI"
BASE COUNT
 347 a 368 c 336 g 316 t
ORIGIN
 Query Match 13.7%; Score 187; DB 12; Length 1367;
 Best Local Similarity 55.0%; Pred. No. 1.4e-37;
 Matches 392; Conservative 0; Mismatches 315; Indels 6; Gaps 1
 QY 337 tatcagaactcaagaggtatgctcctcaaaagctgtctcaaggaagagaagaagcttccca 396
 DB 457 TACCGGATCCAGAGCCACATATATTACAGTCCCTTGGAAAGAGGCTGCTTCC 516
 QY 397 atagcctatcttctgtgtctccacaagaatgcaattatggttgaagagcttaccatgct 456
 DB 517 CTGGCTTACATCATGTGTGATCCACAAAGACTTGATACCTTTGAAGAGGCTTCAAGGCC 576
 QY 457 atatacaaccagacatatattactgcatcattatgatatgataaggaacctgatacttc 516
 DB 577 ATTTCATGCCCCAGAGACGCTCTACTGTGTGCATGTGGATAGTAAGGGAGACACCTTC 636
 QY 517 aaagtgtccctgaacaattagctaaagtgtcttcccaatatcttcattgttcccaaatla 576
 DB 637 AAAGAAGCGGTGGCGCAGTACTACGTGTTTCCCAATGCTTCCGCGGCTCTAAGGTG 696

OY	577	gagctgtgggaataatgcacatttccagaactcgaaggctgattaattgcttgctgcac	636
Db	697	GAGCAGGtGGtCTATNGtTNGtGGCTTCTCTCGCTCCAGGCTGTCTGAACtTCATtGAAGAAR	756
OY	637	cttcctgaagtcttcaatccatggaatatgtttcaactctgtgtggccaagattccc	696
Db	757	CTGGtGGCTCCtCAAGTGtCCCTGGAAAtATtGCTCTCAACACtCGCGGGCAGGACTtTCCr	816
OY	697	ctgaagtcacaatttgaatttgtctcagatgtgtcaaaaactcaatgtggaacaattgtg	756
Db	817	TtGAAAACCACAAGGAATAATTAAcCATtGTGAAGAAttTAAGGGCAAAAAAAtCATCr	876
OY	757	gagacgtgtgaaccccccaaacagtaaatgtgaagaattccatcacatgtgaacttaga	816
Db	877	cCGGAgAGtGtCCCCCTGCCTAcATAGtTGTtAGCGAtCTAAAtTTGTACtCCAGAACG	936
OY	817	cgggtgccttatgatargtgaagctaacaaataagacaacacatctccaagaaagacc	876
Db	937	AAAGCCAAtAGAGOGAtATTtTATGc-----ATAAACAAAtATTtTGAAGActCCACCr	990
OY	877	ccccataacattcagatatgttgttgcaagtgctattctgtttttaagcaagcatgtgt	936
Db	991	CCACACCAAtCGANtCATtTACTTtGGCACACCCtTACCTGGCCtCACCCGGGActTTGTG	1050
OY	937	aatatattttcaacaactcatctgttcaagaacttttttcctgtgtctaagaacacctac	996
Db	1051	AACtTtATtCCGAtNGANGAAGGCCAtTGTCTCTtTAGAGtGTtTAAGAtACTAr	1110
OY	997	tctcctgtatggcaacttttgggtccacttgattcggtgttcacagaataacttg	1049
Db	1111	AGCCtGTAWGAACAATTtTtGGTGACACtCAAtAGGAtTCCAGAGtGtCCCTG	1163
RESULT	9		
AKO08234			
LOCUS	AKO08234	1862 bp	mRNA
DEFINITION	Mus musculus adult male small intestine cDNA, RIKEN full-length		
ACCESSION	enriched library, clone:201001JH22, full insert sequence.		
VERSION	AKO08234.1		
KEYWORDS	AKO08234.1 GI:12842295		
SOURCE	CAP trapper.		
	Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to		
	mRNA, clone:201001JH22.		
	clone:201001JH22.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
ADTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
JOURNAL	1 (bases 1 to 1862)		
PUBMED	Carninci,P. and Hayashizaki,Y.		
REFERENCE	High-efficiency full-length cDNA cloning		
ADTHORS	Methods in enzymology. 303, 19-44 (1999)		
JOURNAL	99279253		
PUBMED	10349636		
REFERENCE	2 (bases 1 to 1862)		
ADTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,		
JOURNAL	Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.		
PUBMED	Normalization and subtraction of cap-trapper-selected cDNAs to		
REFERENCE	prepare full-length cDNA libraries for rapid discovery of new genes		
ADTHORS	Genome research. 10 (10), 1617-1630 (2000)		
JOURNAL	20499374		
PUBMED	11042159		
REFERENCE	3 (bases 1 to 1862)		
ADTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,		
JOURNAL	Kono,H., Akiyama,Y., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,		
PUBMED	Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishize,T., Harada,A.,		
REFERENCE	Yamanoto,R., Matsumoto,H., Sakaueuchi,S., Ikegami,T., Kashiwagi,K.,		
ADTHORS	Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,		
JOURNAL	Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,		
PUBMED	Okazaki,Y., Muramatsu,M., Inoue,Y., Kiru,A. and Hayashizaki,Y.		
REFERENCE	RIKEN integrated sequence analysis (RISA) system-384-format		
ADTHORS	sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)		

MEDLINE
PUBMED
11076861

20530913

REFERENCE
AUTHORS

JOURNAL
REFERENCES
AUTHORS

TITLE
JOURNAL
REFERENCES
AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FAYOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 1862)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Anon,H., Arai,A.,
Arikawa,T., Carinici,P., Fukuda,S., Fukushima,Y., Fununo,M.,
Hasegami,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Horl,F.,
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawaji,J.,
Kojima,Y., Konno,H., Kouda,M., Koys,S., Kurihara,C., Matsuyama,T.,
Miyazaki,Y., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y.,
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
Sasak,i,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sorabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takashashi,F.,
Tanaka,T., Tajima,Y., Toyota,T., Yamamura,T., Yasunishi,A.,
Yoshida,K., Yoshino.M., Muramatsu,M. and Hayashizaki.Y.

Direct Submission
Submitted (10-JUL-2000) yoshihide hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory In RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAACAACGCCATTCCTTTTTTTTGTNN 3'] . cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 20.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'
GAGAGAAGATTGGCAGTTAAATAATCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

location/Qualifiers

1..1862

/organism="Mus musculus"

/strain="CS7BL/6j"

/db_xref="taxon:10090"

/db_xref="MGD:MGI:1904808"

/db_xref="MGD:MGI:1919337"

/clone="2010013H22"

/sex="male"

/tissue-type="small intestine"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

240..1436

/note="putative"

/codon_start=1

/protein_id="BAB25548.1"

/db_xref="GI:12842296"

/translation="MDIDSEFQSYOVCGRDLTYTKRLPAKSINSGVINGECKAVNO
ALLNNEDEKKKOQLFTPADYLIRMTADENHETKRFPLOYPLSKREASFPALYSNYVE
KIENEELKLNCVITPVONVCCHMDQSSPEPKCAVRIVISCFEPVFIAISKVSVAVAS
WSROADLNMEDLLQSPPVMKYILNTCGDFLPRTAKNAVAKALKUKGNMSKEVP
PPHKSRMKYHYEVTDLTLMHTSKRRTPPNPLLTFGTANAYVASRPDIIEFVNASKAR
OLEWVDITYSPDELHWALTQRSMPWGSDPLHKFEPMDSMRALPARLTMYDHGDILE
NGAYSTCSGHORAVCVSGSDHLWLIIHNHHLLANKFDKVDDNVLCGLEELYLRHRA
LYGEFL"

BASE COUNT

ORIGIN

484 a 463 C 478 g 437 t

Query Match 12.8%; Score 174.2; DB 12; Length 1862;

Best Local Similarity 51.5%; Pred. No. 3e-34;
Matches 506; Conservative 0; Mismatches 438; Indels 39; Gaps 3;

OY	316	atgccagatattgtgacatttataagacccaagagtgatgctcaaaagctgtgctca	375
Db	438	ATGACACAGACTGTGGACACTTCAMACCAAGAGAAATTATACAGGTCCACTGAGC	497
OY	376	aaggaggaganaaagcttcccaatagccatcttcttgtgtgtgcacaaagatgcaattg	435
Db	498	AAGAAAGGCGCACCTTCCCATGTGCTACTCCATGCTGGTGCATGAGAAATGGAAC	557
OY	436	gttaaaagcgtacacatgctataataaacagacaaatttaccgtaacccattgat	495
Db	558	TTCCAAAGGTTCGCGAGCTGTACACCCCTCAGAAATATATCTGTGCTCACATGAT	617
OY	496	cgtaaggacccgtgatacccttaaaagttgcacatgaaacaatttgaagtgtcttccaat	555
Db	618	CAGAAGCTTTCAGAACCCCTTAAACAGAGCGATCGAGGCCATCGTGCATCTTCCCAAT	677
OY	556	atttcattgttcccaattagaagcgtgtgaaatgtcccaattccaaagccagagct	615
Db	678	GCTTCATAGCTAGTAAGTGTGTCTCAGTGTGTCTATCTCTCTGTGTCCAGGATGCAAGCT	737
OY	616	gatttaaattgtctgtcggaacctctgaaagtctcaatccagttggaatatgtatcaac	675
Db	738	GACCTAAACTCATGGAAGATCTCTTCAGAGCCCCCTGCATGGAATACCTCTGTAAC	797
OY	676	tgtgtgtggcaagattttccctgaagtcaaatlttgaattgtgtgtaaggttgaaaaa	735
Db	798	ACCTGTGGGACAGCTTCCATPAAACACATCTGATGATGGTCAAGGCCCTCAAGCTA	857
OY	736	ctcaatggaagcaaatatgtgtgagacggttgaaccccccaagtaaatgtgaagaatc	795
Db	858	TTGAAAGGGCAGAACAGATAGAGTCAAGAGTACCACCTCCACATTAATAATCCCGCTGG	917
OY	796	acttaccatctgaacttaagcgggtgtgcctatgtaatatgtgaaagtcaccaataagaca	855
Db	918	AAATATCACTTATGAGGTGACA-----GACACATTGCACATATGAC	956
OY	856	aacatctccaaggaagacccccccaatacaatcagaatattgtgtgagtgctatatt	915
Db	957	AGCAAGAGGAAGAGCGCCGCCACTAATATACCTAACATGTTCACCTGGAAATGCCCTACATG	1010
OY	916	gttttaagttaagcatttgtttaatatatttttaacaactlccatcgttcaagcctttt	975
Db	1017	GTGGCTTCTCGAGACTTATTGAACAGCTGTTCAGTAACTCAAAAGCCCGGCAACTATTC	1077
OY	976	gocctgtctaaagacacatctccctgatatagacacttttggcttaccctgtatcgggt	1033
Db	1077	GAGTGGTAATAAAGAACACCTAATATATGCCGATGAGCACCCTTGGCCACCCCTCCACGCGCC	1133
OY	1036	ccaagataaactgtg---ggagaattccaagatcagcccaagatgtgtgtcatctgcaagat	1095
Db	1137	TCTGTGAGTCCCTGGATWCAGATCCCTTGATCGAAATTTTGACATGTACAGACATGAGACC	1193
OY	1093	aagactgcgctgtcaaatgtgaattacatcatgaagcctttt-----ctat	1133
Db	1197	ATTGGAGACATACCAAGATGGTACGACCATGAGGAGACATTTGAAGACGGGCGACCTTAC	1253
OY	1138	cccaattgtactgtgattccaccttcgaaaggtgtgtatttatgtgaagctgcgaattgaag	1115
Db	1257	ACGTTCTTCCTCAGGATCCACACAGGGGCTCTGTGTGTTATGGGTACAGGGGACTCTCAC	1313
OY	1198	tggctatcaaaagatggaacatgtgttctataaaatttgaatttcaagttggaccctatc	1225
Db	1317	TGGATACCTTCAGAACCATACACTCTTTGGCCAAAGTTGACCCCAAGGTGATGATAAT	1373
OY	1258	ttgatataatgtcttggcagaaa	1280
Db	1377	GTTCCTCAGTGTTTAGAAAGAA	1399

[illegible]

	BASE COUNT	ORIGIN
a	218	others
c	158	t
g	219	
t	246	

Query Match	Similarity	Score	DB	Length
Best Local	54.48%	Pred. No.28-33		
Matches	371;	Conservative	10;	Mismatches 289; Indels 12; Gaps 2
Qy	345	tctaaaggttatgtcccaaaagctgtctccaagsgaggaagagcttcccaatagccta	404	
Db	767	TGGAAACCCTATGTAAACAATAAACTCTGTGAAGAAGGSGTGGTCCCTTTCCTTA	708	
Qy	405	ttccttggttgtccacaagaatgcaattatgtgtgaaggtcatccatgcatatacaa	464	
Db	707	CACAGGACCATTCACAAAAGACTTCGGCACTTTGGAGAGCTTTTAAAGCGATTATAT	648	
Qy	465	ccagacaatatattactgcataccatattatcgtlaagagacctgaatcttcaaaigtgc	524	
Db	647	GCOCCTAAATGTCTACTGTGTGTGACCTGGATCGAAGGGGAGCGAGATCCCTTTAAAGTGC	588	
Qy	525	catgaacaatttagctaagtgtcttcccaatatatttcattgtctccaatataaggtctg	584	
Db	587	AGTAAACAGTCATCTAGCTGTCTCCCAATGCTTTCTGGCTTCCAAARAANKAAGTCGGT	528	
Qy	585	ggaataatgccacaatttccagactccagagctgtaattaaattgtgttcgagaccttga	644	
Db	527	TGCTATGGGGKKAATCTCCADGCTCCAMGCTGACGCGAATGCTGTGAAGACTTGCC	468	
Qy	645	gttctcaatccagttgaataatagttaaccaactgttgggcaagatttcccttgaaatc	704	
Db	467	CTCTGAAGTCCCTGTAAGATATTCTATCAACACCTCGGGCAACAACTTTCCCTGAAAC	408	
Qy	705	aaatttgatgtgtgtcgaggttgaanaaaccaatgagcaaatatgtttgagaacgtc	764	
Db	407	CACACAGGAAATGTCTCAGATCTRAAGGATTTTAAAGGAAATAATACACCCCGGATT	348	
Qy	765	gaaaccccccaacagtaaatgtgaaagatcactta----catcatgaacttgaacggt	821	

Db	347	GCCTCCCTCCGACACACAGCTGTGGAGGACAAATATCTTCCACCAAGAATTCTTAACCA	288
Qy	822	gcctatgaatatgtgagcttaccaatgaagaaacatctccaagaaacacccccca	861
Db	287	CAAAATTTCTAGCTG-----ATTAAACAATAATTTAAATACTCGCTCCGCA	237
Qy	882	taacatcagatatgtgtggcagtgctatttgttitaagtcacgatitgttaata	941
Db	236	TGAATGTTGATTTACTTTTSSCAGGCGCTACGCTGCTCACAAAGGACTTTGGCTAACTT	177
Qy	942	tatttcaacaacatcactcgttcaagaacttttgcctgtgttaagaacatatctcc	1001
Db	176	CGTCTCCAAAGACAGAGCTGCACATTACTTACTCTCTGTCACAKAGACCTTAAGCC	117
Qy	1002	tgatgacgacttttggctacc	1023
Db	116	CGACGAACATTTCTGCGTGACC	95

RESULT	11
CNS04UIM/c	
LOCUS	
DEFINITION	CNS04UIM 970 bp DNA GSS 24 MAY-2000
	Tetradodon nigroviridis genome survey sequence T3
	042KX9 of library A from Tetradodon nigroviridis, genomic survey
	sequence.
ACCESSION	AL307831
VERSION	AL307831.1
KEYWORDS	GI:8213194
SOURCE	GSS; genome survey sequence.
ORGANISM	Tetradodon nigroviridis.
	Tetradodon nigroviridis

REFERENCE

HORS

OT E

IRNAI.

HORS

CONCLUSION

HORS

RNAI

Page 2

2

Mr.

ches

204

Db	519	GTGCAACGTTAATTGTCCCGGATATTCGAAATGACCCGATGAGGTGGGAAGTCTTT	460
QY	264	ggaataaagaagaagsgaacatcattgaacttggagatgaatgatgtgtggaatgaccag	323
Db	459	GNTCAATCCGAGAAACCTCTGCCCGGAGTGGACGAAAGCCTGACCAACTGACTTN	400
QY	324	tgaattgcacattatcgaactcaagaagttagtgcacaagaactgtgtccaaagagga	383
Db	399	TGACTCGAAGAGCTTTGTTCAGGGCCAGAGTTACGACAAAGAGTCNTTGGAAATGCA	340
QY	384	gaaagactcccaataagactctcttctgtgttcacaaagaatgcataatgatgttgaag	443
Db	339	GCGAGANNITTCCTCGGGGTACTAAGTGTGTGGCAANATATGGGTGATGTGGAGG	280
QY	444	gcttatccatgctatatatacaacagacaaatattactgcataccattatgactgtaagc	503
Db	279	GCCTCCAGAGGCGGTGACTCCGCCCATATACATCTACTGCTCCACTCGACCTGAGCTG	220

Y 504 acctgagaccctcctcaagccttgccagaagcaatttagctaaagtgactctcccaattatccat
 Db 219 ccccttaccagttcattnttcggccatagaggccctggcgttcgcttccgccaacncttcat
 QY 564 tgccttccaattatagagcgtcgtgataatgcccacatttccaaactccagagctgatttaa
 Db 159 cgggtccaaagcggaggggtggttcatatcagggggcttcacgcggctgaanaagccactca
 QY 624 ttgccttcgcagaccttcgaagcttccaaatccagctgaaatttgtatcaactctgtg
 Db 99 ctctcctctgcgaccttcctgaagctcagagctcagaagtgaaagtgaactcattccctcgg
 QY 684 gcaagatttccctcctgaagtcocaaatttgaattgtgtc 722
 Db 39 ccaggatttttccctcctcaggtccaaacatcgnnnnntncc 1

RESULT	12
CNSOLUWZ/c	
LOCUS	756 bp DNA GSS
CNSOLUWZ	
DEFINITION	Tetradon nitroviridis genome survey sequence PUC-ORI end of c
	12-MAY-2000

ACCESSION	AL168380
VERSION	AL168380.1
KEYWORDS	GI:7806437
	GSS; genome survey sequence.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes
Tetraodontidae; Tetraodon

TITLE	NAME
Characterization and repeat analysis of the compact genome of the	Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

REFERENCE
AUTHORS
2 (bases 1 to 756)
Roest-Crollius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, B., A., Fizames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Motrocobab, I.

JOURNAL
 Unpublished
 3 (bases 1 to 756)
 REFERENCE
 Genoscope.
 AUTHORS

COMMENT This sequence is a single read and was generated as part of a scale clone-end sequencing project of the *Heterodon nigriviridis* genome. For more information, please take a look at

FEATURES

Source Location/Qualifiers
 1.756
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG197CC03SP1-end : puc-ori"
 BASE COUNT 169 a 189 c 224 g 166 t 8 others
 ORIGIN

Query Match 11.6%; Score 157.6; DB 13; Length 756;
 Best Local Similarity 61.2%; Pred. No. 5e-30;
 Matches 267; Conservative 2; Mismatches 166; Indels 1; Gaps 1;

QY 848 taaggaacaacatctccaagaagacaccccccataacattcagatattgttgcagtg 907
 DB 754 TCAAAACACAGCGTGGCCAAAGCCCGCCCGCCGACATCCAGKCTTCATCGGCAAGC 695
 QY 908 ctatttcttttaagcaagcatttgtaataatatttcacaacactcattcgtcaag 967
 DB 694 CTTATTGCTGTGTGCGGGACTTTTGGCAGACGTCGAACAGAGGAGCTGGCCAGG 635
 QY 968 acttttgcctgtctaaagacacatactctcctgatagacatttgggttaacctga 1027
 DB 634 ACTTCCTGGCTGGTCTGCTGACACACTGCTGCGAGACACTCTGGGCGCACCTGG 575
 QY 1028 ttgggttccaggaatactctgagagatttccagatcagccagagatgtctgattgc 1087
 DB 574 TCAGGCTCCCGGGGTCCCGCCCAATCCCGCTCCCAAGCGGATGACAGATTTGA 515
 QY 1088 agagtaagactcgccttgcagatggaattactatgaaggcttttctatccagttga 1147
 DB 514 GAAGTAAGACCGCGGTGTGAATGGAATCTGTGAGAGGAGGAGAGCTGACCCGCTGCA 455
 QY 1148 ctggtctcacccttcgaagctgtgattatgtgagctgcgaattaaagtggttaca 1207
 DB 454 CGGGACACACCTCGCGAGCGGTGACATCTACGGGCGCGGACTT-CGCTGCTGCTCG 396
 QY 1208 aagaagacattgtttctataaattgatttctaagtgtagacctatctgttaaat 1267
 DB 395 GCTTGGAACACTGTTCCCAAGAGTTGACCCCAAGTGAGCCCGGTTCTGTACAGT 336
 QY 1268 gcttgacagaagaact 1283
 DB 335 GTTTGAGAGAGAGCT 320

RESULT 13
 BE005007/c 526 bp mRNA EST 05-JUN-2000
 LOCUS BE005007
 DEFINITION MR0-BN0115-020300-001-803 BN0115 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE005007.1 GI:8265240
 VERSION BE005007.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 526)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.R.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel.: +55-11-7704922
 Fax: +55-11-7707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-MR0-BN0115-020
 300-001-a036r3-2000-03-02&tl=1)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 526.
 Location/Qualifiers
 1.526
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BN0115"
 /dev_stage="Adult"
 /note="Organ: breast, normal; Vector: puc18; Site: 1; SmaI;
 Site: 2; SmaI; A mini-library was made by cloning products
 derived from ORSTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 126 a 125 c 136 g 139 t
 ORIGIN

Query Match 10.9%; Score 148.8; DB 10; Length 526;
 Best Local Similarity 60.3%; Pred. No. 8.7e-28;
 Matches 246; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 345 tctaagaggtatgctcaaaagcttctcaagaagaggaagaaagcttccaatgctca 404
 DB 497 TCAAGCCACTATATGTAACGAAACACTCTGTGAAGAGGCTGGTTCCTTTCCTTA 438
 QY 405 ttcttgggtgttcacaaagatgcaattatgttgaaggttatccatgctataca 464
 DB 437 CACAGTACCATTCACAAAGACTTGGCACTTTTGAGCGGCTTCAGGCGCATTTAT 378
 QY 465 ccagcaacaattatcagatcaccatgatcgtaagacactgataactcaagttgc 524
 DB 377 GCCCCCAAGTGTACTGTGTCACCTGTGACAGAGGCGAGCGATGCTTTAAAGTGC 318
 QY 525 catgaacaatttagcagatgcttcccaattatccattgcttccaaattagagctgt 584
 DB 317 AGTAAACACTTACTACGCTGCTCCCAATGCTTTTGTGCTCCAAAGAGAGTCGT 258
 QY 585 ggaatagccacattccagactcagactcagactttaaattgcttgcagactctgaa 644
 DB 257 TGTATATGGGGGATCTCCAGGCTCCAGCTGACCTGAACTGCTGGAAGACTTGTGC 198
 QY 645 gtcttcaatccagtggaatatttatacaactggttgggcaagatttccctgaagtc 704
 DB 197 CTCTGAAGTTCCTGGAGATGATGATCAACACCCCGGCGCAAGACTTCCCTGAAAC 138

RESULT 14
 BE292814 612 bp mRNA EST 13-JUL-2000
 LOCUS BE292814
 DEFINITION 601105179F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:298090 5',
 mRNA sequence.
 ACCESSION BE292814
 VERSION BE292814.1 GI:9175472
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 612)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
AUTHORS
NIH-MGC <http://mgc.nci.nih.gov/>
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at: image.llnl.gov
Plate: LHCW6 row: m column: 19
High quality sequence stop: 606.

FEATURES

source
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOB7; Site: 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAG(c). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)".

BASE COUNT

181 a 137 c 140 g 154 t

ORIGIN

Query Match 10.3%; Score 148.8; DB 10; Length 612;
Best Local Similarity 54.0%; Pred. No. 8,9e-28;
Matches 340; Conservative 0; Mismatches 272; Indels 18; Gaps 1;

395 caataagcctatcttctggtgtcccaagaatgcaattatggttgaagagctatccatg 454
1 CTATTGATACCTATATGGTATCATGAGAGATGGAAGACTTTGAAAGGCTACTCGAG 60
455 ctataacacccagcaaatattctatgcatcatcattatgctgaagcaccctgatacc 514
61 CTGTGATGGCCCTCAGACATATACGTGTCCATGTGAGAGTCCGCCAGAACTT 120
515 tcaagtgatgaatgaacattagctgaaggtcttcccaatttctatgcttccaat 574
121 TCAAGAGGCGGCTCAAGCAATATTCTTCTCCCAATATGCTTCATAGCCAGTAAG 180
575 tagaggtgtggaatagccacattccagactccagagctggttaaatgtcttcg 634
181 TGGTTCGGGTGTTATGCTCTCGTCCAGGGTGCAGGTGCACTTCAATGATGAGAG 240
635 acccttgaaagcttcaatccagctggaatatttcaactgtgtggtggaagatttc 694
241 ACTTGCTCCAGAGCTCGCTGCAATATCTTCTGATGATGATGATGATGATGATG 300
695 cccctgaagctcaatttgaattgtgtcagagttgaaataacatcgaatgagcaaatgt 754
301 CTATTAAGAGCAATGCGAGATGTTCCAGGCTCTCAAGATGTTGAATGGAGAGATAGA 360
755 tgaagcgtgtgaaaccccaacagtaaatgtgaaagatcacttccatcgaacta 814
361 TGAAGTCAGAGTACCTCTCAAGCAAGAACCCGTGGAATATCACTTTGAGTAG 420
815 gaaggtgcttatgatatgtgaagctaccataaggaataacatccatcgaagagc 874
421 TGAAG-----ACACATTACACTTACCAACGAAGAGATCTTC 462
875 ccccccataacattcagatatgtgtgagctgcttattgttttaagtaagcatttg 934

Db 463 CCCTATATATTTACTATGTTTACAGGGAATGCTTACATGTGCTCCGAGATTTCG 522
Qy 935 ttaatatatttcaacaacatccatcgttcaagacttttgcgtgtcctaagacac 994
Db 523 TCCACATGTTTGAAGAACCTTAATCCCAACACTGATGATGATGATGATGATGATG 582
Qy 995 actcctgatgagcacttgggtacct 1024
Db 583 ATAGCCCAATGATACACCTCTGGCCACT 612

RESULT 15
BG185546 873 bp mRNA EST 21-APR-2001
LOCUS
DEFINITION
RST495 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION
BG185546.1 GI:13707233
VERSION
BG185546.1 GI:13707233
KEYWORDS
EST.
SOURCE
ORGANISM
Homo sapiens
human.

REFERENCE
AUTHORS
Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krishnoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Velloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.
and Ducar,M.

TITLE
Creation of Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL
Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com

High quality sequence stop: 454.
Location/Qualifiers

FEATURES

source
1..873
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

Query Match 10.4%; Score 142; DB 11; Length 873;
Best Local Similarity 62.3%; Pred. No. 5,3e-26;
Matches 223; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

345 tctaaaggtatgtctcaaaagctgtctcaagaagagaagacgttcccaatagccta 404
Db 374 TCGAAGCCACTATGTAACGAGAAACACTCTCTGAAAGAGAGCTGGTCCCTTATGCTTA 315
405 tctctgtgtgtccacaagaatgcaatattgtgtgaaagcttaccatcattatataca 464
Db 314 CACAGTGAACCAATCCAAAGACTTGGCACTTTGAGAGGCTTCAAGGGAGATTATAT 255
Qy 465 ccagacacatattctcgtcatcattatgcatcgttaagcaccctgataacctcaagttgc 524
Db 254 GCCCAAAATGCTACTGTTGACACCTGATCAGAGGAGGAGGATGCTTTAAAGGTC 195
Qy 525 catgacaatttagctagcttcccaatatttcttccatattagagagctgt 584

Fri May 3 10:57:54 2002

us-09-645-192-1.rst

Page 12

Db 194 AGTGAACAGTACTCAGCTCTTCCAAAAGCTTTTCGCTCCACGAAGAGAGTGCCT 135

Qy 585 ggaatattcccaacttttccagaccacccagggcggatattaatgctgtgcgaccttcgaa 644

Db 134 TGTCTATGGGGGAGTCTCTCAGGCTCCAGCTGACCTGACCTGGAAACCTTTGGGC 75

Qy 645 gtcttcaatccagtggaatatgtatacctaacttggggggaagatttttccccgaag 702

Db 74 CTCCTAAATTCCCTGAGAGTATGTCTATCAACACTCGGGGCAACACTTTCCCTGTAG 17

```
Search completed: May  2, 2002, 09:55:22
Job time: 5812 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2002, 10:36:36 ; Search time 1980.83 seconds

(without alignments)
11343.305 Million cell updates/sec

Title: us-09-645-192-1

Perfect score: 1362
Sequence: 1 atgaagaatattcaaatgtta.....atctcaactaccacatcatga 1362

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1358776

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_pl:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_cm:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1362	100.0	1362	6	AX087935
2	1362	100.0	3435	2	AF133035
3	1260	92.5	184590	2	AC093259
4	25	1.8	31	6	AX087939
5	21	1.5	21	6	AX087937
6	21	1.5	696	8	CNS01AY6
7	21	1.5	9643	8	CNS01AY6
8	21	1.5	24650	2	AC023731
9	21	1.5	84011	2	AP002094
10	21	1.5	133008	2	AC011375
11	21	1.5	146190	2	AC073826
12	21	1.5	155359	2	AL358777
13	21	1.5	161676	2	AC023461
14	21	1.5	178098	9	HS168L15
15	21	1.5	188353	9	AC009480
16	21	1.5	195277	2	AL592214
17	21	1.5	186788	2	AC021626
18	21	1.5	301051	3	AE003487
19	20	1.5	32	6	AX087938
20	20	1.5	173	9	HUMPSABD
21	20	1.5	185	9	HUMPRBS11
22	20	1.5	478	9	HUMSP508
23	20	1.5	732	10	MUSCXC32
24	20	1.5	1086	14	FLHRZD
25	20	1.5	2165	1	EU059304
26	20	1.5	2705	10	AF327184S1
27	20	1.5	61274	9	AL391815
28	20	1.5	70339	9	AL513123
29	20	1.5	82101	9	AC005694
30	20	1.5	89862	9	AC007164
31	20	1.5	110443	9	AL354692
32	20	1.5	120891	2	AC011285
33	20	1.5	126138	9	HSAR18000
34	20	1.5	156374	9	AC007559
35	20	1.5	162323	2	AL596283
36	20	1.5	163189	2	AC073556
37	20	1.5	170294	2	AC016845
38	20	1.5	173899	9	AC010150
39	20	1.5	174596	2	AC022812
40	20	1.5	174725	9	AC007486
41	20	1.5	175092	2	AC013261
42	20	1.5	181994	2	AC090822
43	20	1.5	183807	2	AC026544
44	20	1.5	183147	2	AC080188
45	20	1.5	185421	2	AC079856
46	20	1.5	198105	2	AC026160
47	20	1.5	226281	9	AC004673
48	20	1.5	318488	9	AC005529
49	19	1.4	496	11	G55641
50	19	1.4	1087	14	AF217223
51	19	1.4	1196	8	HSIRPS05
52	19	1.4	1752	8	D89126
53	19	1.4	2835	8	SCSCY1GEN
54	19	1.4	2857	1	AF278705
55	19	1.4	3065	8	HSZ92867
56	19	1.4	3528	8	SCYGL083W
57	19	1.4	3535	8	YSPCTL
58	19	1.4	11649	1	AE001880
59	19	1.4	33450	2	CEH27104
60	19	1.4	33869	3	CEH38C2
61	19	1.4	36258	3	CEH29F11
62	19	1.4	37194	8	HSV210E9
63	19	1.4	37228	8	SPCC757
64	19	1.4	40600	3	CEH20D1
65	19	1.4	43349	9	HSAC000367
66	19	1.4	47079	2	AC083776
67	19	1.4	51831	2	AL138999
68	19	1.4	52290	8	AC025314
69	19	1.4	66136	8	AB024031
70	19	1.4	70169	8	AF296833

C 71	19	1.4	71302	2	AC087760	AC087760 Homo sapi	C 144	19	1.4	188613	2	AC012133	AC012133 Homo sapi
C 72	19	1.4	78260	2	AC015439	AC015439 Drosophila	C 145	19	1.4	188643	2	AC092422	AC092422 Homo sapi
C 73	19	1.4	83556	2	AC007401	AC007401 Homo sapi	C 146	19	1.4	189032	2	AC025564	AC025564 Homo sapi
C 74	19	1.4	85832	9	HS344F17	HS344F17 Human DNA s	C 147	19	1.4	189602	2	AC023237	AC023237 Homo sapi
C 75	19	1.4	86516	9	AC009493	AC009493 Homo sapi	C 148	19	1.4	190166	33	AC026285	AC026285 Homo sapi
C 76	19	1.4	89490	8	TM15	TM15 Neurospora	C 149	19	1.4	190548	2	AC027677	AC027677 Homo sapi
C 77	19	1.4	89795	8	NC3H10	NC3H10 Neurospora	C 150	19	1.4	190773	9	AC018513	AC018513 Homo sapi
C 78	19	1.4	101270	2	HS483K16	HS483K16 Human DNA	C 151	19	1.4	191570	2	AL591598	AL591598 Homo sapi
C 79	19	1.4	101555	2	AF188026	AF188026 Homo sapi	C 152	19	1.4	192169	2	AC020849	AC020849 Mus muscu
C 80	19	1.4	104810	9	AC003012	AC003012 Human PAC	C 153	19	1.4	192422	2	AC021071	AC021071 Homo sapi
C 81	19	1.4	124833	9	AC007064	AC007064 Homo sapi	C 154	19	1.4	194070	2	AC084313	AC084313 Homo sapi
C 82	19	1.4	125856	9	HS1187M17	HS1187M17 Human DNA	C 155	19	1.4	194734	2	AC011386	AC011386 Homo sapi
C 83	19	1.4	126736	2	AC044788	AC044788 Homo sapi	C 156	19	1.4	196375	2	AC073107	AC073107 Homo sapi
C 84	19	1.4	129517	2	AC016400	AC016400 Homo sapi	C 157	19	1.4	198146	2	AC023499	AC023499 Homo sapi
C 85	19	1.4	130336	2	AC006346	AC006346 Homo sapi	C 158	19	1.4	198748	2	AC016719	AC016719 Homo sapi
C 86	19	1.4	132183	2	AC011376	AC011376 Homo sapi	C 159	19	1.4	199255	2	AC090720	AC090720 Homo sapi
C 87	19	1.4	136168	9	HS1049G11	HS1049G11 Human DNA	C 160	19	1.4	199702	2	AC081178	AC081178 Homo sapi
C 88	19	1.4	139079	2	AC022195	AC022195 Homo sapi	C 161	19	1.4	200149	2	AC006548	AC006548 Homo sapi
C 89	19	1.4	140864	2	AC074355	AC074355 Oryza sat	C 162	19	1.4	203449	2	AC073210	AC073210 Homo sapi
C 90	19	1.4	143094	2	AL591114	AL591114 Homo sapi	C 163	19	1.4	212549	9	AC009492	AC009492 Homo sapi
C 91	19	1.4	143549	2	AP003910	AP003910 Oryza sat	C 164	19	1.4	214631	2	AC060805	AC060805 Homo sapi
C 92	19	1.4	145520	2	AC016036	AC016036 Homo sapi	C 165	19	1.4	218470	1	BSUB0013	BSUB0013 Homo sapi
C 93	19	1.4	147052	2	AC090687	AC090687 Homo sapi	C 166	19	1.4	220735	2	AC087889	AC087889 Homo sapi
C 94	19	1.4	147619	9	AL512358	AL512358 Human DNA	C 167	19	1.4	224076	2	NC084019	NC084019 Mus muscu
C 95	19	1.4	149903	2	AC090510	AC090510 Homo sapi	C 168	19	1.4	282700	1	BACJH642	BACJH642 Homo sapi
C 96	19	1.4	150650	2	AC064845	AC064845 Homo sapi	C 169	19	1.4	317511	3	AE003440	AE003440 Drosophila
C 97	19	1.4	154028	2	HSAC002087	HSAC002087 Human BAC	C 170	19	1.4	342068	6	AX087940	AX087940 Homo sapi
C 98	19	1.4	154561	2	AC087344	AC087344 Homo sapi	C 171	19	1.3	250	20	AY021388	AY021388 Homo sapi
C 99	19	1.4	154811	2	AC024699	AC024699 Homo sapi	C 172	19	1.3	250	11	G15516	G15516 human STS S
C 100	19	1.4	154953	2	AL591023	AL591023 Homo sapi	C 173	19	1.3	297	5	HEFICRCH1	HEFICRCH1 Homo sapi
C 101	19	1.4	155021	9	AC007759	AC007759 Homo sapi	C 174	19	1.3	297	5	HEFICRCH1	HEFICRCH1 Homo sapi
C 102	19	1.4	155026	2	AL442128	AL442128 Human DNA	C 175	19	1.3	297	5	AE06052	AE06052 Homo sapi
C 103	19	1.4	156301	2	AC083900	AC083900 Homo sapi	C 176	19	1.3	592	6	AF098981	AF098981 Homo sapi
C 104	19	1.4	156596	2	AL355479	AL355479 Homo sapi	C 177	19	1.3	599	11	G64409	G64409 B661B4/Sp6
C 105	19	1.4	157340	2	AC015652	AC015652 Homo sapi	C 178	19	1.3	740	11	RCU14374	RCU14374 Rana catesbe
C 106	19	1.4	158391	2	AC008023	AC008023 Homo sapi	C 179	19	1.3	807	6	AG6054	AG6054 Sequence 87
C 107	19	1.4	159517	2	AC018511	AC018511 Homo sapi	C 180	19	1.3	807	6	AG6056	AG6056 Sequence 93
C 108	19	1.4	159743	2	AC011357	AC011357 Homo sapi	C 181	19	1.3	807	6	AG6056	AG6056 Sequence 93
C 109	19	1.4	160274	2	AC091953	AC091953 Homo sapi	C 182	19	1.3	807	6	AG6056	AG6056 Sequence 93
C 110	19	1.4	160422	2	AC011087	AC011087 Homo sapi	C 183	19	1.3	903	6	AE272025	AE272025 Bacillus
C 111	19	1.4	162505	2	AC024258	AC024258 Homo sapi	C 184	19	1.3	919	8	AF166485	AF166485 Glycine m
C 112	19	1.4	162593	2	AC022576	AC022576 Homo sapi	C 185	19	1.3	995	11	CNS06K78	CNS06K78 end of
C 113	19	1.4	162875	9	AC018509	AC018509 Homo sapi	C 186	19	1.3	1047	14	TM9REPPPO	TM9REPPPO end of
C 114	19	1.4	163404	2	AL355308	AL355308 Homo sapi	C 187	19	1.3	1086	5	AF111218	AF111218 Cyprinell
C 115	19	1.4	164405	2	AL513488	AL513488 Homo sapi	C 188	19	1.3	1251	8	HEFICRCH1	HEFICRCH1 Homo sapi
C 116	19	1.4	166919	2	AL591675	AL591675 Mus muscu	C 189	19	1.3	1457	8	AF349521	AF349521 Rattus norveg
C 117	19	1.4	166971	9	AL157812	AL157812 Human DNA	C 190	19	1.3	1755	8	AF327420	AF327420 Rattus norveg
C 118	19	1.4	168025	9	AC024364	AC024364 Homo sapi	C 191	19	1.3	2475	8	EGT082	EGT082 Homo sapi
C 119	19	1.4	168278	2	AC037492	AC037492 Homo sapi	C 192	19	1.3	2745	9	HS252060	HS252060 Homo sapi
C 120	19	1.4	169268	2	AC027674	AC027674 Homo sapi	C 193	19	1.3	2858	10	BC006700	BC006700 Homo sapi
C 121	19	1.4	171490	2	AC068209	AC068209 Homo sapi	C 194	19	1.3	3069	1	SUSVEBIV	SUSVEBIV Homo sapi
C 122	19	1.4	171690	2	AC025925	AC025925 Homo sapi	C 195	19	1.3	3543	6	SAPKNH	SAPKNH Homo sapi
C 123	19	1.4	171764	2	AC022803	AC022803 Homo sapi	C 196	19	1.3	3664	1	AY056399	AY056399 Homo sapi
C 124	19	1.4	172749	2	AC068042	AC068042 Homo sapi	C 197	19	1.3	3958	6	YC05886A	YC05886A Homo sapi
C 125	19	1.4	173268	9	AL138760	AL138760 Human DNA	C 198	19	1.3	4116	3	AB060286	AB060286 Homo sapi
C 126	19	1.4	174437	2	AC007768	AC007768 Homo sapi	C 199	19	1.3	4240	3	BMOXD	BMOXD Homo sapi
C 127	19	1.4	175020	9	AC022317	AC022317 Homo sapi	C 200	19	1.3	4866	8	YC05886A	YC05886A Homo sapi
C 128	19	1.4	175265	2	AC063964	AC063964 Homo sapi	C 201	19	1.3	5313	8	YC05886A	YC05886A Homo sapi
C 129	19	1.4	176630	2	AC010684	AC010684 Homo sapi	C 202	19	1.3	5332	6	HS0802808	HS0802808 Homo sapi
C 130	19	1.4	177407	2	AC009880	AC009880 Homo sapi	C 203	19	1.3	5355	6	ES00078	ES00078 Homo sapi
C 131	19	1.4	178597	2	AC063962	AC063962 Homo sapi	C 204	19	1.3	5355	6	ES00078	ES00078 Homo sapi
C 132	19	1.4	180568	2	AC006482	AC006482 Homo sapi	C 205	19	1.3	5355	6	ES00078	ES00078 Homo sapi
C 133	19	1.4	180707	2	AC024460	AC024460 Homo sapi	C 206	19	1.3	5374	8	SCYBR112C	SCYBR112C Homo sapi
C 134	19	1.4	181902	2	AC020788	AC020788 Homo sapi	C 207	19	1.3	7360	8	HEFICRCH1	HEFICRCH1 Homo sapi
C 135	19	1.4	183938	2	AC023876	AC023876 Homo sapi	C 208	19	1.3	7500	8	RICMTRP	RICMTRP Homo sapi
C 136	19	1.4	184804	2	AC036174	AC036174 Homo sapi	C 209	19	1.3	8237	8	AWU10505	AWU10505 Homo sapi
C 137	19	1.4	185018	2	AL499602	AL499602 Homo sapi	C 210	19	1.3	8949	8	SC00MDLX	SC00MDLX Homo sapi
C 138	19	1.4	185838	9	AC036191	AC036191 Homo sapi	C 211	19	1.3	9971	1	U67577	U67577 Homo sapi
C 139	19	1.4	186298	2	AC007860	AC007860 Homo sapi	C 212	19	1.3	11730	1	AE002480	AE002480 Homo sapi
C 140	19	1.4	186360	2	AL512429	AL512429 Homo sapi	C 213	19	1.3	16781	9	AL356422	AL356422 Homo sapi
C 141	19	1.4	186526	2	AC019306	AC019306 Homo sapi	C 214	19	1.3	17385	8	OSA277468	OSA277468 Homo sapi
C 142	19	1.4	187020	2	AC026614	AC026614 Homo sapi	C 215	19	1.3	17769	9	AL391814	AL391814 Homo sapi
C 143	19	1.4	187894	9	AC007320	AC007320 Homo sapi	C 216	19	1.3	21115	3	AB005911	AB005911 Bombyx mo

C 217	18	1.3	27333	3	CEL128B4	AF026206 Caenorhab	C 290	18	1.3	122537	2	AP003858	AP003858 Oryza sat
C 218	18	1.3	28585	3	U41031	A41031 Caenorhabd1	C 291	18	1.3	123778	2	AC069294	AC069294 Homo sapi
C 219	18	1.3	29605	3	AB020747	AB020747 Arabidops	C 292	18	1.3	124529	2	AP003166	AP003166 Homo sapi
C 220	18	1.3	31537	2	AC091528-4	Continuation (5 of	C 293	18	1.3	125397	2	AC008550	AC008550 Homo sapi
C 221	18	1.3	33649	3	CBRG44J05	AC0864618 Caenorhab	C 294	18	1.3	125595	2	AL139149	AL139149 Homo sapi
C 222	18	1.3	35241	3	CELY39A3B	AC006748 Caenorhab	C 295	18	1.3	126038	8	AP000367	AP000367 Oryza sat
C 223	18	1.3	37526	3	AL139825	AL139825 Human DNA	C 296	18	1.3	126039	8	AP0003187	AP0003187 Oryza sat
C 224	18	1.3	39339	3	CEK08H10	283113 Caenorhabd1	C 297	18	1.3	126839	8	HS19408	HS19408 Homo sapi
C 225	18	1.3	39439	3	AF067219	AF067219 Caenorhab	C 298	18	1.3	127066	2	AC084265	AC084265 Homo sapi
C 226	18	1.3	42503	3	HSU115G11	271187 Human DNA s	C 299	18	1.3	129090	2	AP003931	AP003931 Oryza sat
C 227	18	1.3	43138	3	AL590100	AL590100 Human DNA	C 300	18	1.3	129263	9	AL360089	AL360089 Human DNA
C 228	18	1.3	43526	3	CEIC10H11	UB8311 Caenorhabd1	C 301	18	1.3	129355	9	AC004853	AC004853 Homo sapi
C 229	18	1.3	46239	1	AC027138	AC027138 Staphyloc	C 302	18	1.3	129422	2	AL589677	AL589677 Human DNA
C 230	18	1.3	46848	2	AC013885	AC013885 Drosophila	C 303	18	1.3	131279	2	AC016441	AC016441 Homo sapi
C 231	18	1.3	51730	2	AC018578	AC018578 Drosophila	C 304	18	1.3	131458	9	AC002386	AC002386 Human BAC
C 232	18	1.3	53546	2	AC084346	AC084346 Homo sapi	C 305	18	1.3	133787	9	AC015540	AC015540 Homo sapi
C 233	18	1.3	61226	2	AL139392	AL139392 Human DNA	C 306	18	1.3	134464	10	AC084052	AC084052 Mus Muscu
C 234	18	1.3	63743	2	AC073401	AC073401 Homo sapi	C 307	18	1.3	136129	2	AC011050	AC011050 Homo sapi
C 235	18	1.3	66424	2	AL512509	AL512509 Human DNA	C 308	18	1.3	136672	2	AC092451	AC092451 Homo sapi
C 236	18	1.3	69748	8	SCRACIT	X78993 S.cerevisia	C 309	18	1.3	137032	2	AC007481	AC007481 Homo sapi
C 237	18	1.3	69787	2	AC003984	AC003984 Homo sapi	C 310	18	1.3	137339	2	AC087318	AC087318 Homo sapi
C 238	18	1.3	69961	2	AC008419	AC008419 Homo sapi	C 311	18	1.3	137339	2	AC090683	AC090683 Oryza sat
C 239	18	1.3	71000	8	SEPB8B7	AL032684 S.pombe c	C 312	18	1.3	137659	9	AL139232	AL139232 Homo sapi
C 240	18	1.3	73371	2	AC004872	AC004872 Homo sapi	C 313	18	1.3	137659	9	AP001607	AP001607 Human DNA
C 241	18	1.3	75073	2	AC087648	AC087648 Homo sapi	C 314	18	1.3	137879	8	AP003233	AP003233 Oryza sat
C 242	18	1.3	75073	2	CEY48A6B	AL023844 Caenorhab	C 315	18	1.3	138430	2	AL592204	AL592204 Danio rer
C 243	18	1.3	75935	2	AC022818	AC022818 Homo sapi	C 316	18	1.3	138459	2	AP003936	AP003936 Homo sapi
C 244	18	1.3	76825	2	AL159165	AL159165 Homo sapi	C 317	18	1.3	138633	2	AL356360	AL356360 Homo sapi
C 245	18	1.3	79561	2	AL135799	AL135799 Homo sapi	C 318	18	1.3	139152	6	AP002525	AP002525 Oryza sat
C 246	18	1.3	79874	2	AC027252	AC027252 Homo sapi	C 319	18	1.3	140229	8	AP003143	AP003143 Oryza sat
C 247	18	1.3	81369	6	AX034587	AX034587 Sequence	C 320	18	1.3	140899	2	AC083769	AC083769 Homo sapi
C 248	18	1.3	82456	2	HS153G14	AL031118 Human DNA	C 321	18	1.3	140952	8	AP003046	AP003046 Homo sapi
C 249	18	1.3	84096	2	AC074170	AC074170 Mus muscu	C 322	18	1.3	141432	2	AC015699	AC015699 Homo sapi
C 250	18	1.3	84570	2	AL359552	AL359552 Human DNA	C 323	18	1.3	141475	2	AC002980	AC002980 Homo sapi
C 251	18	1.3	86828	2	AC020569	AC020569 Homo sapi	C 324	18	1.3	141581	2	AC0088979	AC0088979 Homo sapi
C 252	18	1.3	88394	8	HSU734P14	AL043650 Human DNA	C 325	18	1.3	143291	2	AC026414	AC026414 Homo sapi
C 253	18	1.3	90220	2	AP003917	AP003917 Oryza sat	C 326	18	1.3	143710	2	AP002899	AP002899 Oryza sat
C 254	18	1.3	90935	2	AC005168	AC005168 Arabidops	C 327	18	1.3	144314	8	AP000835	AP000835 Homo sapi
C 255	18	1.3	91893	8	AC008802	AC008802 Homo sapi	C 328	18	1.3	144319	9	AC007968	AC007968 Homo sapi
C 256	18	1.3	92388	2	HSJ393D12	AL132776 Human DNA	C 329	18	1.3	144324	2	AC022662	AC022662 Homo sapi
C 257	18	1.3	92817	2	AF257499	AF257499 Homo sapi	C 330	18	1.3	144440	2	AC037436	AC037436 Homo sapi
C 258	18	1.3	93153	2	AP000664	AP000664 Homo sapi	C 331	18	1.3	145242	2	AC013457	AC013457 Homo sapi
C 259	18	1.3	94779	2	AC010068	AC010068 Drosophila	C 332	18	1.3	146118	6	AC069267	AC069267 Homo sapi
C 260	18	1.3	95039	2	AC012816	AC012816 Homo sapi	C 333	18	1.3	146335	2	AP002521	AP002521 Oryza sat
C 261	18	1.3	95546	9	AP003908	AP003908 Oryza sat	C 334	18	1.3	146360	9	HS1156M12	HS1156M12 Homo sapi
C 262	18	1.3	96090	2	AC005868	AC005868 Homo sapi	C 335	18	1.3	146937	2	AL592423	AL592423 Homo sapi
C 263	18	1.3	97656	9	AP000629	AP000629 Homo sapi	C 336	18	1.3	147035	2	AP003418	AP003418 Homo sapi
C 264	18	1.3	97656	9	AP000629	AP000629 Homo sapi	C 337	18	1.3	147038	8	AL133326	AL133326 Human DNA
C 265	18	1.3	99908	8	ATE24G16	AL138647 Arabidops	C 338	18	1.3	147117	8	AC069300	AC069300 Oryza sat
C 266	18	1.3	100906	8	AP003736	AP003736 Oryza sat	C 339	18	1.3	147334	2	AP003349	AP003349 Oryza sat
C 267	18	1.3	102378	2	AC012072	AC012072 Homo sapi	C 340	18	1.3	147352	2	AC018538	AC018538 Homo sapi
C 268	18	1.3	105686	2	AL5924289	AL5924289 Homo sapi	C 341	18	1.3	147492	2	AL358781	AL358781 Homo sapi
C 269	18	1.3	105769	2	AC008520	AC008520 Homo sapi	C 342	18	1.3	147518	33	AC027518	AC027518 Homo sapi
C 270	18	1.3	107101	2	AL356736	AL356736 Homo sapi	C 343	18	1.3	147712	9	AC011299	AC011299 Homo sapi
C 271	18	1.3	108399	2	AC008837	AC008837 Homo sapi	C 344	18	1.3	147741	2	AC092629	AC092629 Homo sapi
C 272	18	1.3	108569	2	AC020813	AC020813 Mus muscu	C 345	18	1.3	147865	2	AL591402	AL591402 Homo sapi
C 273	18	1.3	108634	2	AP000706	AP000706 Homo sapi	C 346	18	1.3	148018	2	AC068861	AC068861 Homo sapi
C 274	18	1.3	109074	2	AC090750-3	Continuation (4 of	C 347	18	1.3	148454	8	AC007023	AC007023 Homo sapi
C 275	18	1.3	110000	2	AC091528-3	Continuation (4 of	C 348	18	1.3	148762	8	AP002843	AP002843 Oryza sat
C 276	18	1.3	110000	2	AL162501-1	Continuation (2 of	C 349	18	1.3	148920	2	AP003343	AP003343 Oryza sat
C 277	18	1.3	110000	2	AL359978-0	Continuation (2 of	C 350	18	1.3	149931	2	AC034112	AC034112 Homo sapi
C 278	18	1.3	110000	2	AL359978-1	Continuation (2 of	C 351	18	1.3	149618	9	AP000556	AP000556 Homo sapi
C 279	18	1.3	110000	2	AL359978-1	Continuation (2 of	C 352	18	1.3	149901	2	AC022243	AC022243 Homo sapi
C 280	18	1.3	110403	2	AC058813	AC058813 Homo sapi	C 353	18	1.3	150036	8	AP000557	AP000557 Homo sapi
C 281	18	1.3	110952	9	AL136369	AL136369 Human DNA	C 354	18	1.3	150150	8	AP002559	AP002559 Oryza sat
C 282	18	1.3	111182	9	AC073608	AC073608 Homo sapi	C 355	18	1.3	150379	8	AP003074	AP003074 Oryza sat
C 283	18	1.3	113366	9	AL591804	AL591804 Human DNA	C 356	18	1.3	150519	8	AP002993	AP002993 Lotus Jap
C 284	18	1.3	112884	9	AL136966	AL136966 Homo sapi	C 357	18	1.3	150583	9	AC026884	AC026884 Homo sapi
C 285	18	1.3	112902	9	AC005230	AC005230 Homo sapi	C 358	18	1.3	150849	2	AP003541	AP003541 Oryza sat
C 286	18	1.3	113515	9	HSJ944F13	AL117353 Human DNA	C 359	18	1.3	150849	2	AC020890	AC020890 Homo sapi
C 287	18	1.3	117557	9	HSJ1004I9	AL121912 Human DNA	C 360	18	1.3	151321	2	AC012276	AC012276 Homo sapi
C 288	18	1.3	118653	9	AP003738	AP003738 Oryza sat	C 361	18	1.3	151335	2	AC023302	AC023302 Homo sapi
C 289	18	1.3	119972	2	AP004029	AP004029 Oryza sat	C 362	18	1.3	151340	2	AC069132	AC069132 Homo sapi

C 363	18	1.3	151755	2	AC013299	AC013299 Homo sapi	C 436	18	1.3	165262	2	AP003181	AP003181 Homo sapi
C 364	18	1.3	152327	2	AC060759	AC060759 Homo sapi	C 437	18	1.3	165289	2	HS5915	AL442639 Homo sapi
C 365	18	1.3	152582	2	AC027014	AC027014 Homo sapi	C 438	18	1.3	165388	2	AL590966	AL590966 Homo sapi
C 366	18	1.3	152935	2	AC012642	AC012642 Homo sapi	C 439	18	1.3	165388	9	AP000961	AP000961 Homo sapi
C 367	18	1.3	152959	2	AC011275	AC011275 Homo sapi	C 440	18	1.3	165915	2	AC092922	AC092922 Homo sapi
C 368	18	1.3	153215	2	HS24A17	AL035452 Homo sapi	C 441	18	1.3	166071	9	AC022336	AC022336 Homo sapi
C 369	18	1.3	153289	2	AC013553	AC013553 Homo sapi	C 442	18	1.3	166287	2	AC026030	AC026030 Homo sapi
C 370	18	1.3	153408	2	AC011995	AC011995 Homo sapi	C 443	18	1.3	166620	2	AC020516	AC020516 Homo sapi
C 371	18	1.3	153764	9	AL355350	AL355350 Homo sapi	C 444	18	1.3	166782	2	AC021516	AC021516 Homo sapi
C 372	18	1.3	154091	9	AP003819	AP003819 Homo sapi	C 445	18	1.3	167025	2	AC021380	AC021380 Homo sapi
C 373	18	1.3	154172	9	AC079386	AC079386 Homo sapi	C 446	18	1.3	167163	2	AC021705	AC021705 Homo sapi
C 374	18	1.3	154619	1	D90917	D90917 Synechocyst	C 447	18	1.3	167437	2	AC023068	AC023068 Homo sapi
C 375	18	1.3	155278	9	AC040161	AC040161 Homo sapi	C 448	18	1.3	167631	9	AL158829	AL158829 Homo sapi
C 376	18	1.3	155576	2	AC055712	AC055712 Homo sapi	C 449	18	1.3	167862	9	AC011966	AC011966 Homo sapi
C 377	18	1.3	155664	2	AC073140	AC073140 Homo sapi	C 450	18	1.3	168293	2	AL451137	AL451137 Homo sapi
C 378	18	1.3	155774	2	AL592202	AL592202 Homo sapi	C 451	18	1.3	168334	2	AC027208	AC027208 Homo sapi
C 379	18	1.3	155804	2	AC019241	AC019241 Homo sapi	C 452	18	1.3	168467	2	AP001143	AP001143 Homo sapi
C 380	18	1.3	156002	2	AC036110	AC036110 Homo sapi	C 453	18	1.3	168538	2	AC091722	AC091722 Homo sapi
C 381	18	1.3	156069	2	AC079256	AC079256 Homo sapi	C 454	18	1.3	168561	2	AC091989	AC091989 Homo sapi
C 382	18	1.3	156334	2	AC025755	AC025755 Homo sapi	C 455	18	1.3	168623	2	AC084013	AC084013 Homo sapi
C 383	18	1.3	156403	2	AC078823	AC078823 Homo sapi	C 456	18	1.3	168624	33	AC021874	AC021874 Homo sapi
C 384	18	1.3	156457	2	AC093014	AC093014 Homo sapi	C 457	18	1.3	168720	2	AC092987	AC092987 Homo sapi
C 385	18	1.3	156457	9	AC004383	AC004383 Homo sapi	C 458	18	1.3	168777	2	AL356773	AL356773 Homo sapi
C 386	18	1.3	156485	2	AL359701	AL359701 Homo sapi	C 459	18	1.3	168893	2	AC032018	AC032018 Homo sapi
C 387	18	1.3	156724	2	AP001630	AP001630 Homo sapi	C 460	18	1.3	168990	2	AC079324	AC079324 Homo sapi
C 388	18	1.3	156957	2	AP001977	AP001977 Homo sapi	C 461	18	1.3	169237	9	AC009516	AC009516 Homo sapi
C 389	18	1.3	157074	2	AC055708	AC055708 Homo sapi	C 462	18	1.3	169337	2	AL356462	AL356462 Homo sapi
C 390	18	1.3	157077	2	AC022504	AC022504 Homo sapi	C 463	18	1.3	169401	2	AC059056	AC059056 Homo sapi
C 391	18	1.3	157086	2	AP000552	AP000552 Homo sapi	C 464	18	1.3	169939	2	AC079765	AC079765 Homo sapi
C 392	18	1.3	157183	9	AC011731	AC011731 Homo sapi	C 465	18	1.3	170224	2	AL139341	AL139341 Homo sapi
C 393	18	1.3	157358	2	AC024906	AC024906 Homo sapi	C 466	18	1.3	170508	2	AC022450	AC022450 Homo sapi
C 394	18	1.3	157358	2	AC021972	AC021972 Homo sapi	C 467	18	1.3	170602	2	AC022536	AC022536 Homo sapi
C 395	18	1.3	158026	2	AL158213	AL158213 Homo sapi	C 468	18	1.3	170796	2	AC069416	AC069416 Homo sapi
C 396	18	1.3	158039	2	AC021840	AC021840 Homo sapi	C 469	18	1.3	170939	2	AL354684	AL354684 Homo sapi
C 397	18	1.3	158148	2	AC015463	AC015463 Homo sapi	C 470	18	1.3	170939	2	AC069536	AC069536 Homo sapi
C 398	18	1.3	158236	9	AC011631	AC011631 Homo sapi	C 471	18	1.3	170995	2	AC092040	AC092040 Homo sapi
C 399	18	1.3	158723	8	AP002863	AP002863 Homo sapi	C 472	18	1.3	171044	2	AC092787	AC092787 Homo sapi
C 400	18	1.3	159100	33	AC021023	AC021023 Homo sapi	C 473	18	1.3	171073	2	AC073380	AC073380 Homo sapi
C 401	18	1.3	159217	9	AC016950	AC016950 Homo sapi	C 474	18	1.3	171185	2	AL356507	AL356507 Homo sapi
C 402	18	1.3	159222	2	AC092419	AC092419 Homo sapi	C 475	18	1.3	171206	2	AC026573	AC026573 Homo sapi
C 403	18	1.3	159287	2	AC079119	AC079119 Homo sapi	C 476	18	1.3	171245	9	AC044791	AC044791 Homo sapi
C 404	18	1.3	159387	2	AC010162	AC010162 Homo sapi	C 477	18	1.3	171490	33	AC036166	AC036166 Homo sapi
C 405	18	1.3	159453	2	AL596267	AL596267 Homo sapi	C 478	18	1.3	171523	2	AC010206	AC010206 Homo sapi
C 406	18	1.3	159586	2	AF377946	AF377946 Homo sapi	C 479	18	1.3	171621	2	AC011961	AC011961 Homo sapi
C 407	18	1.3	159586	2	AF377946	AF377946 Homo sapi	C 480	18	1.3	171828	2	AC022247	AC022247 Homo sapi
C 408	18	1.3	159637	2	AC007990	AC007990 Homo sapi	C 481	18	1.3	172571	2	AC027484	AC027484 Homo sapi
C 409	18	1.3	159816	2	AC011075	AC011075 Homo sapi	C 482	18	1.3	172747	9	AL359835	AL359835 Homo sapi
C 410	18	1.3	160196	2	AC016701	AC016701 Homo sapi	C 483	18	1.3	172942	9	AC062031	AC062031 Homo sapi
C 411	18	1.3	160421	2	AC025005	AC025005 Homo sapi	C 484	18	1.3	173141	2	AC027508	AC027508 Homo sapi
C 412	18	1.3	160492	2	AL512584	AL512584 Homo sapi	C 485	18	1.3	173445	2	AL162380	AL162380 Homo sapi
C 413	18	1.3	160541	2	AC007944	AC007944 Homo sapi	C 486	18	1.3	173465	2	AC084374	AC084374 Homo sapi
C 414	18	1.3	160760	2	AC079772	AC079772 Homo sapi	C 487	18	1.3	173506	2	AC016323	AC016323 Homo sapi
C 415	18	1.3	160953	2	AC018407	AC018407 Homo sapi	C 488	18	1.3	173521	33	AC067850	AC067850 Homo sapi
C 416	18	1.3	161394	2	AC010759	AC010759 Homo sapi	C 489	18	1.3	173560	2	AL513262	AL513262 Homo sapi
C 417	18	1.3	161459	2	AC023594	AC023594 Homo sapi	C 490	18	1.3	173764	2	AC012464	AC012464 Homo sapi
C 418	18	1.3	161687	2	AC048351	AC048351 Homo sapi	C 491	18	1.3	173879	2	AC092600	AC092600 Homo sapi
C 419	18	1.3	161876	2	AC073634	AC073634 Homo sapi	C 492	18	1.3	173891	33	AC009545	AC009545 Homo sapi
C 420	18	1.3	162194	2	AC026592	AC026592 Homo sapi	C 493	18	1.3	173987	2	AC027165	AC027165 Homo sapi
C 421	18	1.3	162346	9	AC004551	AC004551 Homo sapi	C 494	18	1.3	174095	2	AL356788	AL356788 Homo sapi
C 422	18	1.3	162361	2	AC023779	AC023779 Homo sapi	C 495	18	1.3	174129	2	AC092617	AC092617 Homo sapi
C 423	18	1.3	162609	9	AC021171	AC021171 Homo sapi	C 496	18	1.3	174165	33	AL391480	AL391480 Homo sapi
C 424	18	1.3	162868	9	AC010348	AC010348 Homo sapi	C 497	18	1.3	174465	2	AC022582	AC022582 Homo sapi
C 425	18	1.3	163147	2	AC024445	AC024445 Homo sapi	C 498	18	1.3	174771	2	AC015454	AC015454 Homo sapi
C 426	18	1.3	163345	2	AL360002	AL360002 Homo sapi	C 499	18	1.3	174782	9	AC079150	AC079150 Homo sapi
C 427	18	1.3	163464	2	AC007462	AC007462 Homo sapi	C 500	18	1.3	174832	9	AF280107	AF280107 Homo sapi
C 428	18	1.3	163612	2	AC001794	AC001794 Homo sapi	C 501	18	1.3	175093	2	AC027185	AC027185 Homo sapi
C 429	18	1.3	163624	2	AC009780	AC009780 Homo sapi	C 502	18	1.3	175184	9	AC022367	AC022367 Homo sapi
C 430	18	1.3	164232	2	AC079064	AC079064 Homo sapi	C 503	18	1.3	175332	2	AL583808	AL583808 Homo sapi
C 431	18	1.3	164518	2	AC068210	AC068210 Homo sapi	C 504	18	1.3	175387	2	AC087616	AC087616 Homo sapi
C 432	18	1.3	165130	2	AL596113	AL596113 Homo sapi	C 505	18	1.3	175627	2	AC092664	AC092664 Homo sapi
C 433	18	1.3	165146	9	AL116444	AL116444 Homo sapi	C 506	18	1.3	175670	2	AL358535	AL358535 Homo sapi
C 434	18	1.3	165228	2	CNS01DWS	AL138479 Human sapi	C 507	18	1.3	175902	9	AP000904	AP000904 Homo sapi
C 435	18	1.3	165243	2	AL162408	AL162408 Homo sapi	C 508	18	1.3	175940	9	HS0152L7	HS0152L7 Homo sapi

C 509	18	1.3	176051	9	AC023490	AC023490	Human sapi	582	18	1.3	188043	2	AL500526	AL500526	Human sapi
C 510	18	1.3	176056	9	AL136359	AL136359	Human sapi	583	18	1.3	188152	2	AC025847	AC025847	Human sapi
C 511	18	1.3	176132	2	AC093026	AC093026	Human sapi	584	18	1.3	188180	2	AC068021	AC068021	Human sapi
C 512	18	1.3	176379	2	AC060833	AC060833	Human sapi	585	18	1.3	188358	2	AC009922	AC009922	Human sapi
C 513	18	1.3	176556	2	AL356980	AL356980	Human sapi	586	18	1.3	188420	2	AC025205	AC025205	Human sapi
C 514	18	1.3	176639	9	AC016987	AC016987	Human sapi	587	18	1.3	188966	2	AC020747	AC020747	Human sapi
C 515	18	1.3	176645	9	AC068712	AC068712	Human sapi	588	18	1.3	189385	2	CNS06C7Y	CNS06C7Y	Human chr
C 516	18	1.3	176678	2	AC036213	AC036213	Human sapi	589	18	1.3	189814	2	AC010754	AC010754	Human sapi
C 517	18	1.3	176714	9	AC005618	AC005618	Human sapi	590	18	1.3	190000	2	AC006425	AC006425	Human sapi
C 518	18	1.3	176724	2	AC091882	AC091882	Human sapi	591	18	1.3	190220	2	AC068343	AC068343	Human sapi
C 519	18	1.3	176783	2	AL139281	AL139281	Human sapi	592	18	1.3	190277	2	AC079474	AC079474	Human sapi
C 520	18	1.3	177307	2	AP003524	AP003524	Cyza sat	593	18	1.3	190746	2	AC087191	AC087191	Human sapi
C 521	18	1.3	177696	2	AL139815	AL139815	Human sapi	594	18	1.3	190984	2	AC048332	AC048332	Human sapi
C 522	18	1.3	177716	2	AP001027	AP001027	Human sapi	595	18	1.3	191141	2	AC069374	AC069374	Human sapi
C 523	18	1.3	177717	9	AC061958	AC061958	Human sapi	596	18	1.3	191244	2	AC024222	AC024222	Human sapi
C 524	18	1.3	177992	2	AC040902	AC040902	Human sapi	597	18	1.3	191676	2	AC011844	AC011844	Human sapi
C 525	18	1.3	178373	2	AL357273	AL357273	Human sapi	598	18	1.3	192021	2	AP003536	AP003536	Human sapi
C 526	18	1.3	178401	2	AL357770	AL357770	Human sapi	599	18	1.3	192150	2	AP003025	AP003025	Human sapi
C 527	18	1.3	178414	2	AC025341	AC025341	Human sapi	600	18	1.3	192718	2	AC090366	AC090366	Human sapi
C 528	18	1.3	178585	2	AC092927	AC092927	Human sapi	601	18	1.3	193026	2	AC010158	AC010158	Human sapi
C 529	18	1.3	178714	3	AC022925	AC022925	Human sapi	602	18	1.3	193078	2	AC010907	AC010907	Human sapi
C 530	18	1.3	178796	2	AC018857	AC018857	Human sapi	603	18	1.3	193112	2	AP002811	AP002811	Human sapi
C 531	18	1.3	179026	2	AC018857	AC018857	Human sapi	604	18	1.3	193131	2	AL390766	AL390766	Human sapi
C 532	18	1.3	179134	2	AC022817	AC022817	Human sapi	605	18	1.3	194256	2	AC019158	AC019158	Human sapi
C 533	18	1.3	179369	2	AP000842	AP000842	Human sapi	606	18	1.3	194750	2	AC092961	AC092961	Human sapi
C 534	18	1.3	179396	2	AC017056	AC017056	Human sapi	607	18	1.3	194874	2	AC080090	AC080090	Human sapi
C 535	18	1.3	179460	2	AC012512	AC012512	Human sapi	608	18	1.3	195058	2	AP003087	AP003087	Human sapi
C 536	18	1.3	179556	2	AC018734	AC018734	Human sapi	609	18	1.3	195121	2	AC021187	AC021187	Human sapi
C 537	18	1.3	179622	2	AC092519	AC092519	Felis cat	610	18	1.3	195343	2	AC068342	AC068342	Human sapi
C 538	18	1.3	179833	9	AL162233	AL162233	Human sapi	611	18	1.3	195343	2	AC018338	AC018338	Human sapi
C 539	18	1.3	179912	9	AL133415	AL133415	Human sapi	612	18	1.3	195472	2	AL357336	AL357336	Human sapi
C 540	18	1.3	179947	2	AL591853	AL591853	Human sapi	613	18	1.3	195913	2	AC026603	AC026603	Human sapi
C 541	18	1.3	179971	2	AC013395	AC013395	Human sapi	614	18	1.3	196267	2	AC005386	AC005386	Human sapi
C 542	18	1.3	180072	2	AC017092	AC017092	Human sapi	615	18	1.3	196361	9	AC019235	AC019235	Human sapi
C 543	18	1.3	180129	2	AC021026	AC021026	Human sapi	616	18	1.3	196680	9	AC018694	AC018694	Human sapi
C 544	18	1.3	180499	3	AC007817	AC007817	Drosophila	617	18	1.3	197031	2	AC021155	AC021155	Human sapi
C 545	18	1.3	180865	2	AP003766	AP003766	Oryza sat	618	18	1.3	197547	2	AC013494	AC013494	Human sapi
C 546	18	1.3	181137	2	AC084303	AC084303	Human sapi	619	18	1.3	198239	2	AL358073	AL358073	Human sapi
C 547	18	1.3	181175	2	AL158214	AL158214	Human sapi	620	18	1.3	198239	2	AL358073	AL358073	Human sapi
C 548	18	1.3	181264	2	AC016962	AC016962	Human sapi	621	18	1.3	198978	2	AC011312	AC011312	Human sapi
C 549	18	1.3	181305	2	AC013415	AC013415	Human sapi	622	18	1.3	200337	9	AL590608	AL590608	Human sapi
C 550	18	1.3	181899	2	AC015904	AC015904	Human sapi	623	18	1.3	201027	2	AC040896	AC040896	Human sapi
C 551	18	1.3	182133	2	AC016235	AC016235	Human sapi	624	18	1.3	201035	2	AC016744	AC016744	Human sapi
C 552	18	1.3	182268	2	AC015591	AC015591	Human sapi	625	18	1.3	203076	2	AC087061	AC087061	Human sapi
C 553	18	1.3	182274	2	AC023435	AC023435	Human sapi	626	18	1.3	203111	2	AP001647	AP001647	Human sapi
C 554	18	1.3	183203	9	AL356750	AL356750	Human sapi	627	18	1.3	203241	2	AC008282	AC008282	Human sapi
C 555	18	1.3	183204	2	AC092625	AC092625	Human sapi	628	18	1.3	203405	2	AC042964	AC042964	Human sapi
C 556	18	1.3	183289	2	AP002829	AP002829	Human sapi	629	18	1.3	203407	2	AC006174	AC006174	Human sapi
C 557	18	1.3	183475	2	AC023487	AC023487	Human sapi	630	18	1.3	204154	2	AL450352	AL450352	Human sapi
C 558	18	1.3	183512	2	AC027552	AC027552	Human sapi	631	18	1.3	204220	2	AC022938	AC022938	Human sapi
C 559	18	1.3	183756	8	AC007789	AC007789	Oryza sat	632	18	1.3	204310	2	AC017094	AC017094	Human sapi
C 560	18	1.3	183818	2	AP001202	AP001202	Human sapi	633	18	1.3	206056	2	AC004387	AC004387	Human sapi
C 561	18	1.3	183833	2	AC011085	AC011085	Human sapi	634	18	1.3	206388	2	AC013698	AC013698	Human sapi
C 562	18	1.3	183873	2	AC068044	AC068044	Human sapi	635	18	1.3	206457	9	AC013268	AC013268	Human sapi
C 563	18	1.3	183970	2	AC027002	AC027002	Human sapi	636	18	1.3	206925	2	AC080008	AC080008	Human sapi
C 564	18	1.3	183980	2	AC079014	AC079014	Human sapi	637	18	1.3	208326	2	AC022888	AC022888	Human sapi
C 565	18	1.3	184226	2	AL157786	AL157786	Human sapi	638	18	1.3	208326	2	AP003530	AP003530	Human sapi
C 566	18	1.3	184342	9	AL353766	AL353766	Human sapi	639	18	1.3	208765	2	AC010736	AC010736	Human sapi
C 567	18	1.3	183542	9	AL445473	AL445473	Human sapi	640	18	1.3	208765	2	AC008778	AC008778	Human sapi
C 568	18	1.3	183549	9	AL161646	AL161646	Human sapi	641	18	1.3	210202	2	AL365497	AL365497	Human sapi
C 569	18	1.3	183933	2	AC017022	AC017022	Human sapi	642	18	1.3	210814	2	HS0543C6	HS0543C6	Human sapi
C 570	18	1.3	183970	2	AC017022	AC017022	Human sapi	643	18	1.3	211025	2	AL580875	AL580875	Human sapi
C 571	18	1.3	184010	2	AC016856	AC016856	Human sapi	644	18	1.3	212110	2	AL581707	AL581707	Human sapi
C 572	18	1.3	184226	2	AL157786	AL157786	Human sapi	645	18	1.3	212656	9	AC007957	AC007957	Human sapi
C 573	18	1.3	185229	2	AP001173	AP001173	Human sapi	646	18	1.3	212656	9	AC007957	AC007957	Human sapi
C 574	18	1.3	185229	10	AC006507	AC006507	Mus muscu	647	18	1.3	214222	2	AC074140	AC074140	Human sapi
C 575	18	1.3	185755	2	AC092146	AC092146	Human sapi	648	18	1.3	214445	2	AL365497	AL365497	Human sapi
C 576	18	1.3	185880	2	AC079377	AC079377	Mus muscu	649	18	1.3	216441	2	AC027331	AC027331	Human sapi
C 577	18	1.3	185939	2	AP002829	AP002829	Human sapi	650	18	1.3	216431	2	AL365477	AL365477	Human sapi
C 578	18	1.3	186414	2	AC012454	AC012454	Human sapi	651	18	1.3	218675	2	AC030227	AC030227	Human sapi
C 579	18	1.3	186572	2	AC079130	AC079130	Mus muscu	652	18	1.3	218894	2	AC073152	AC073152	Human sapi
C 580	18	1.3	186896	2	AC037486	AC037486	Human sapi	653	18	1.3	219564	2	AL390207	AL390207	Human sapi
C 581	18	1.3	187044	2	AL158164	AL158164	Human sapi	654	18	1.3	219635	2	AL390207	AL390207	Human sapi
C 581	18	1.3	187445	2	AC024639	AC024639	Human sapi	654	18	1.3	221437	9	AC034242	AC034242	Human sapi

C 655	18	1.3	222477	2	AP001847	AP001847 Homo sapi	C 728	17	1.2	959	2	AC082528	AC082528 Giardlia 1
C 656	18	1.3	224058	9	AC008513	AC008513 Homo sapi	C 729	17	1.2	967	2	AC067107	AC067107 Giardlia 1
C 657	18	1.3	224400	2	AC008758	AC008758 Homo sapi	C 730	17	1.2	981	14	AB018697	AB018697 Human rot
C 658	18	1.3	224614	3	AE003761	AE003761 Drosophila	C 731	17	1.2	984	14	GA026376	U26376 Group A rot
C 659	18	1.3	227949	10	AC090431	AC090431 Mus muscu	C 732	17	1.2	981	14	AY002485	AY002485 HIV-1 iso
C 660	18	1.3	228239	2	AC011405	AC011405 Homo sapi	C 733	17	1.2	995	2	AC065417	AC065417 Giardlia 1
C 661	18	1.3	235586	2	AC024173	AC024173 Mus muscu	C 734	17	1.2	996	2	AC036912	AC036912 Giardlia 1
C 662	18	1.3	240825	6	AX087869	AX087869 Sequence	C 735	17	1.2	1001	2	AC054114	AC054114 Giardlia 1
C 663	18	1.3	254871	2	AL586136	AL586136 Mus muscu	C 736	17	1.2	1001	9	AE279968	AE279968 Homo sapi
C 664	18	1.3	257728	2	AC006846	AC006846 Caenorhab	C 737	17	1.2	1001	9	AE279969	AE279969 Homo sapi
C 665	18	1.3	258369	2	AC091771	AC091771 Mus muscu	C 738	17	1.2	1001	9	AE279970	AE279970 Homo sapi
C 666	18	1.3	258654	2	AC008579	AC008579 Homo sapi	C 739	17	1.2	1001	9	AE279971	AE279971 Homo sapi
C 667	18	1.3	260464	2	AC006800	AC006800 Caenorhab	C 740	17	1.2	1001	9	AE279972	AE279972 Homo sapi
C 668	18	1.3	273658	2	AC079423	AC079423 Mus muscu	C 741	17	1.2	1001	9	AE279973	AE279973 Homo sapi
C 669	18	1.3	281423	2	AL139335	AL139335 Homo sapi	C 742	17	1.2	1001	9	AE279974	AE279974 Homo sapi
C 670	18	1.3	282895	2	AL133478	AL133478 Homo sapi	C 743	17	1.2	1001	9	AE279975	AE279975 Homo sapi
C 671	18	1.3	287571	2	AC092358	AC092358 Homo sapi	C 744	17	1.2	1001	9	AE279976	AE279976 Homo sapi
C 672	18	1.3	294754	2	AC055741	AC055741 Homo sapi	C 745	17	1.2	1001	9	AE279977	AE279977 Homo sapi
C 673	18	1.3	297201	2	AC091956	AC091956 Homo sapi	C 746	17	1.2	1001	9	AE279978	AE279978 Homo sapi
C 674	18	1.3	299050	1	AP003131	AP003131 Staphyloc	C 747	17	1.2	1001	9	AE279979	AE279979 Homo sapi
C 675	18	1.3	299719	2	AC006858	AC006858 Caenorhab	C 748	17	1.2	1001	9	AE279980	AE279980 Homo sapi
C 676	18	1.3	300829	2	AL359532	AL359532 Homo sapi	C 749	17	1.2	1001	9	AE279981	AE279981 Homo sapi
C 677	18	1.3	303750	1	AP003133	AP003133 Staphyloc	C 750	17	1.2	1001	9	AE279982	AE279982 Homo sapi
C 678	18	1.3	305622	2	AC006746	AC006746 Caenorhab	C 751	17	1.2	1001	9	AE279984	AE279984 Homo sapi
C 679	18	1.3	319012	3	AE003556	AE003556 Drosophila	C 752	17	1.2	1001	9	AE279985	AE279985 Homo sapi
C 680	18	1.3	322774	3	CEV73F8A	AL132862 Caenorhab	C 753	17	1.2	1001	9	AE279986	AE279986 Homo sapi
C 681	18	1.3	329861	1	NMA522491	AL162756 Neisseria	C 754	17	1.2	1001	9	AE279987	AE279987 Homo sapi
C 682	18	1.3	340000	9	AP001688	AP001688 Homo sapi	C 755	17	1.2	1001	9	AE279988	AE279988 Homo sapi
C 683	18	1.3	340000	9	AP001700	AP001700 Homo sapi	C 756	17	1.2	1001	9	AE279989	AE279989 Homo sapi
C 684	18	1.3	343590	1	AP003359	AP003359 Staphyloc	C 757	17	1.2	1001	9	AE279990	AE279990 Homo sapi
C 685	18	1.3	347225	1	AP003361	AP003361 Staphyloc	C 758	17	1.2	1001	9	AE279992	AE279992 Homo sapi
C 686	18	1.3	349600	6	AX044032	AX044032 Sequence	C 759	17	1.2	1001	9	AE279993	AE279993 Homo sapi
C 687	17	1.2	235	10	MM009498	U09498 Mus musculu	C 760	17	1.2	1001	9	AE279994	AE279994 Homo sapi
C 688	17	1.2	285	10	AY011829	AY011829 Dipodomys	C 761	17	1.2	1001	9	AE279995	AE279995 Homo sapi
C 689	17	1.2	294	6	AX12157	AX12157 Sequence	C 762	17	1.2	1001	9	AE279996	AE279996 Homo sapi
C 690	17	1.2	304	11	DM106D9T	Z83507 D. melanog	C 763	17	1.2	1001	9	AE279997	AE279997 Homo sapi
C 691	17	1.2	341	3	AF021066	AF021066 Achenom	C 764	17	1.2	1001	9	AE279998	AE279998 Homo sapi
C 692	17	1.2	376	11	G48787	G48787 SHGC-83643	C 765	17	1.2	1001	9	AE279999	AE279999 Homo sapi
C 693	17	1.2	381	5	AF118421	AF118421 Pagrus au	C 766	17	1.2	1001	9	AE280001	AE280001 Homo sapi
C 694	17	1.2	399	14	RGR288006	AJ288006 Human gro	C 767	17	1.2	1001	9	AE280002	AE280002 Homo sapi
C 695	17	1.2	488	11	G45513	G45513 SHGC-68699	C 768	17	1.2	1001	9	AE280003	AE280003 Homo sapi
C 696	17	1.2	498	4	HGCL02	V004722 Halichoerus	C 769	17	1.2	1001	9	AE280004	AE280004 Homo sapi
C 697	17	1.2	517	9	HSPA23C5	Z79131 H. sapiens f	C 770	17	1.2	1001	9	AE280005	AE280005 Homo sapi
C 698	17	1.2	629	10	AF303453	AF303453 Mus muscu	C 771	17	1.2	1001	9	AE279983	AE279983 Homo sapi
C 699	17	1.2	631	10	AF303454	AF303454 Mus muscu	C 772	17	1.2	1002	9	AE279991	AE279991 Homo sapi
C 700	17	1.2	638	8	CNAJ2958	AJ002958 Cicer arti	C 773	17	1.2	1002	9	AE280000	AE280000 Homo sapi
C 701	17	1.2	635	10	AY011332	AY011332 Muscardin	C 774	17	1.2	1002	9	AE280007	AE280007 Homo sapi
C 702	17	1.2	653	14	RVGA041	Z80274 Rotavirus g	C 775	17	1.2	1002	9	AE280008	AE280008 Homo sapi
C 703	17	1.2	673	10	AF303451	AF303451 Mus muscu	C 776	17	1.2	1002	9	AC057137	AC057137 Giardlia 1
C 704	17	1.2	687	11	G48874	G48874 SHGC-77728	C 777	17	1.2	1014	2	PRVC95X	PRVC95X
C 705	17	1.2	699	5	TAU77178	U77178 Triticus al	C 778	17	1.2	1047	14	PRVDP7Y	PRVDP7Y
C 706	17	1.2	732	6	AX122913	AX122913 Sequence	C 779	17	1.2	1059	8	AB045607	AB045607 Homo sapi
C 707	17	1.2	733	6	HSAT5594	AJ005594 Homo sapi	C 780	17	1.2	1059	8	AB045607	AB045607 Homo sapi
C 708	17	1.2	784	10	AF303450	AF303450 Mus muscu	C 781	17	1.2	1062	14	KOBYP7A	KOBYP7A
C 709	17	1.2	805	1	AF179592	AF179592 Yldrio ch	C 782	17	1.2	1083	3	AB001513	AB001513 Carabus m
C 710	17	1.2	816	6	A80952	A80952 Sequence 4	C 783	17	1.2	1083	3	AB001514	AB001514 Carabus m
C 711	17	1.2	816	6	A95331	A95331 Sequence 4	C 784	17	1.2	1083	3	AB050677	AB050677 Aristoc
C 712	17	1.2	834	6	A85688	A85688 Sequence 34	C 785	17	1.2	1083	3	AB050687	AB050687 Carabus v
C 713	17	1.2	834	6	AR155181	AR155181 Sequence	C 786	17	1.2	1142	5	AF261254	AF261254 Mantidact
C 714	17	1.2	834	6	E65706	E65706 Genome DNA	C 787	17	1.2	1152	4	AF095465	AF095465 Oryctolag
C 715	17	1.2	842	8	AB024274	AB024274 Citrus un	C 788	17	1.2	1158	10	RNLCKTP	RNLCKTP
C 716	17	1.2	854	8	AY027357	AY027357 Arabidops	C 789	17	1.2	1341	6	AX143113	AX143113 Sequence
C 717	17	1.2	854	8	A85702	A85702 Sequence 36	C 790	17	1.2	1409	14	FLANURAMA	FLANURAMA
C 718	17	1.2	855	6	AR155195	AR155195 Sequence	C 791	17	1.2	1409	14	FLANURAMB	FLANURAMB
C 719	17	1.2	855	6	E65720	E65720 Genome DNA	C 792	17	1.2	1409	14	FLANURAMC	FLANURAMC
C 720	17	1.2	863	6	AR155889	AR155889 Sequence 54	C 793	17	1.2	1413	14	FLANMK	FLANMK
C 721	17	1.2	863	6	AR155382	AR155382 Sequence	C 794	17	1.2	1433	8	ECOPCNLEU	ECOPCNLEU
C 722	17	1.2	863	6	E65907	E65907 Genome DNA	C 795	17	1.2	1431	8	AF170870	AF170870 Saccharom
C 723	17	1.2	911	2	AC053308	AC053308 Giardlia 1	C 796	17	1.2	1458	9	AF177201	AF177201 Homo sapi
C 724	17	1.2	915	13	AF265273	AF265273 Unculture	C 797	17	1.2	1459	10	UBAJ3455	UBAJ3455 Unidentif
C 725	17	1.2	917	8	AF094426	AF094426 Oxyphilis	C 798	17	1.2	1488	10	AF244361	AF244361 Mus muscu
C 726	17	1.2	924	2	AC078530	AC078530 Giardlia 1	C 799	17	1.2	1490	1	AF182102	AF182102 Escherich
C 727	17	1.2	938	2	AC065874	AC065874 Giardlia 1	C 800	17	1.2	1490	1	AF182104	AF182104 Escherich

C 801	17	1.2	1490	1	AF182105	AF182105 Escherich	874	17	1.2	5642	2	AC091129	AC091129 Mus muscu
C 802	17	1.2	1490	1	AF182106	AF182106 Escherich	875	17	1.2	5678	2	AC017339	AC017339 Drosophila
C 803	17	1.2	1490	1	AF182107	AF182107 Escherich	876	17	1.2	5694	3	AC006687	AC006687 Caenorhab
C 804	17	1.2	1498	1	AF182103	AF182103 Escherich	877	17	1.2	5775	5	AB034701	AB034701 Xenopus l
C 805	17	1.2	1507	6	A43711	A43711 Sequence 8	878	17	1.2	5874	9	AF117907	AF117907 Homo sapi
C 806	17	1.2	1507	6	AR106145	AR106145 Sequence	879	17	1.2	5895	1	AF242211	AF242211 Homo sapi
C 807	17	1.2	1539	10	AF217173	AF217173 Mus muscu	880	17	1.2	6329	2	AC019778	AC019778 Drosophila
C 808	17	1.2	1539	9	AF142992	AF142992 Homo sapi	881	17	1.2	6394	9	HSM802644	HSM802644 Homo sapi
C 809	17	1.2	1593	9	AF184344	AF184344 Homo sapi	882	17	1.2	6447	3	DM040214	DM040214 Drosophila
C 810	17	1.2	1594	9	HSU94703	HSU94703 Homo sapien	883	17	1.2	6511	10	AF261919	AF261919 Mus muscu
C 811	17	1.2	1600	9	BC009194	BC009194 Homo sapi	884	17	1.2	6693	9	AF172078	AF172078 Homo sapi
C 812	17	1.2	1612	1	ECORPOS	D17549 Escherichia	885	17	1.2	7299	1	AF242209	AF242209 Escherich
C 813	17	1.2	1642	1	ECOSTGMA38	D13548 Escherichia	886	17	1.2	7350	6	AR145630	AR145630 Sequence
C 814	17	1.2	1686	10	MM094700	U94700 Mus musculu	887	17	1.2	7350	6	II4735	II4735 Sequence 14
C 815	17	1.2	1705	5	AF128489	AF128489 Calotes v	888	17	1.2	7408	10	AF223080	AF223080 Rattus no
C 816	17	1.2	1770	5	XELXADA	L17888 Xenopus lae	889	17	1.2	7679	8	SCYNR071C	SCYNR071C S.cerevisia
C 817	17	1.2	1818	6	AX109977	AX109977 Sequence	890	17	1.2	7850	1	AF242208	AF242208 Escherich
C 818	17	1.2	1846	9	AK026831	AK026831 Homo sapi	891	17	1.2	7850	1	AF242210	AF242210 Escherich
C 819	17	1.2	1863	9	AK027744	AK027744 Homo sapi	892	17	1.2	8462	10	AF060246	AF060246 Mus muscu
C 820	17	1.2	1898	10	MMMY5	X56182 M.musculus	893	17	1.2	8700	8	AF068574	AF068574 Arabidops
C 821	17	1.2	1959	1	AF182950	AF182950 Homo sapi	894	17	1.2	8877	1	BACMCPTLP	BACMCPTLP Arabidops
C 822	17	1.2	2045	1	AF275947	AF275947 Escherich	895	17	1.2	9010	14	AF178440	AF178440 Tritoma
C 823	17	1.2	2051	5	XL077639	U77639 Xenopus lae	896	17	1.2	9557	9	AL390767	AL390767 Human DNA
C 824	17	1.2	2129	3	AF336057	AF336057 Pristina	897	17	1.2	9973	1	AE004295	AE004295 Vibrio ch
C 825	17	1.2	2240	10	MMIAP1B	AB015497 Passiflor	898	17	1.2	10008	1	AE001241	AE001241 Treponema
C 826	17	1.2	2295	8	AB015497	AB015497 Mouse DNA f	899	17	1.2	10098	1	AE006255	AE006255 Lactococc
C 827	17	1.2	2340	10	AF021836	AF021836 Mus muscu	900	17	1.2	10116	1	AE005326	AE005326 Escherich
C 828	17	1.2	2603	9	AK027702	AK027702 Homo sapi	901	17	1.2	10268	1	AE005361	AE005361 Escherich
C 829	17	1.2	2631	14	PHVCGAA	X16672 Pepper Huas	902	17	1.2	10271	1	AE000922	AE000922 Methanoba
C 830	17	1.2	2646	8	AF141928	AF141928 Pelargonol	903	17	1.2	10495	3	EHCP5	EHCP5 Escherich
C 831	17	1.2	2654	8	GMU02013	U20213 Glycine max	904	17	1.2	10573	1	AE005502	AE005502 E.nistolyti
C 832	17	1.2	2654	8	SCYBL097W	X25858 S.cerevisia	905	17	1.2	10580	1	AE001784	AE001784 Thermocog
C 833	17	1.2	2689	8	LESBP1	X89929 L.esculentu	906	17	1.2	10674	14	AF100465	AF100465 Dendro vi
C 834	17	1.2	2729	8	HSM801458	AL133604 Homo sapi	907	17	1.2	10674	14	AF100467	AF100467 Dendro vi
C 835	17	1.2	2751	8	DE70034	DE70034 Muccor javan	908	17	1.2	10674	14	AF100468	AF100468 Dendro vi
C 836	17	1.2	2825	1	AF021260	AF021260 Botreila	909	17	1.2	10674	14	AF100469	AF100469 Dendro vi
C 837	17	1.2	2928	8	AB027507	AB027507 Arabidops	910	17	1.2	10703	1	AE006380	AE006380 Lactococc
C 838	17	1.2	2938	8	BC007291	BC007291 Homo sapi	911	17	1.2	11060	1	AE007595	AE007595 Methanococc
C 839	17	1.2	2949	6	AX099364	AX099364 Sequence	912	17	1.2	11093	1	AE007754	AE007754 Clostridii
C 840	17	1.2	3006	9	AB056773	AB056773 Macaca fa	913	17	1.2	11115	1	AE001228	AE001228 Treponema
C 841	17	1.2	3012	8	CORPOGEM	X6384 O.orienti rpic	914	17	1.2	11129	1	AE000358	AE000358 Escherich
C 842	17	1.2	3323	8	SCYNR072W	U71687 S.cerevisia	915	17	1.2	11457	1	AE006586	AE006586 Streptococ
C 843	17	1.2	3346	5	AF103901	AF103901 Parca fla	916	17	1.2	11507	1	AE003124	AE003124 Bacillus
C 844	17	1.2	3386	6	AX144813	AX144813 Sequence	917	17	1.2	11562	1	AE000242	AE000242 Escherich
C 845	17	1.2	3386	6	AX144813	AX144813 Sequence	918	17	1.2	11663	1	AE008029	AE008029 Agrobact
C 846	17	1.2	3393	5	HOMCSPB	U03072 Human serin	919	17	1.2	12339	1	AE008029	AE008029 Agrobact
C 847	17	1.2	3429	9	AF143003	AF143003 Parca fla	920	17	1.2	12343	1	AE005198	AE005198 Escherich
C 848	17	1.2	3499	9	HSU43519	U13519 Human dyster	921	17	1.2	12349	1	AE005198	AE005198 Escherich
C 849	17	1.2	3588	9	HSM802481	AL157434 Homo sapi	922	17	1.2	13083	1	AE007434	AE007434 Streptococ
C 850	17	1.2	3850	6	DB4439	DB4439 Schistosomach	923	17	1.2	13086	2	AC019563	AC019563 Drosophila
C 851	17	1.2	3949	8	AY033318	AY033318 Uncultured	924	17	1.2	13316	1	U67602	U67602 Methanococc
C 852	17	1.2	4126	1	AF270497	AF270497 Escherich	925	17	1.2	14763	1	AE000734	AE000734 Agulfox a
C 853	17	1.2	4160	6	AX086127	AX086127 Sequence	926	17	1.2	14925	10	MMCLCNV14	MMCLCNV14 Mus muscu
C 854	17	1.2	4160	6	HSM801597	AL136624 Homo sapi	927	17	1.2	15251	1	D90786	D90786 E.coli geno
C 855	17	1.2	4164	1	MBFPDHC	MB4798 Methanobact	928	17	1.2	15401	8	AC023838	AC023838 Arabidops
C 856	17	1.2	4169	1	AFU33282	U33282 Arababaca PC	929	17	1.2	15401	8	AP000385	AP000385 Arabidops
C 857	17	1.2	4343	8	SCYNL087W	U71363 S.cerevisia	930	17	1.2	15892	3	AE001419	AE001419 Plasmodiu
C 858	17	1.2	4363	8	NCUI7251	U17251 Neurospora	931	17	1.2	16019	1	D90785	D90785 E.coli geno
C 859	17	1.2	4421	9	HSM802480	AL157433 Homo sapi	932	17	1.2	17933	6	SCORESDNA	SCORESDNA S.cerevisia
C 860	17	1.2	4470	3	AF041423	AF041423 Plasmodiu	933	17	1.2	18510	2	CE1724C12	CE1724C12 Drosophila
C 861	17	1.2	4470	3	AF041424	AF041424 Plasmodiu	934	17	1.2	18510	2	CE1724C12	CE1724C12 Drosophila
C 862	17	1.2	4499	3	AF041422	AF041422 Plasmodiu	935	17	1.2	19049	2	AC020853	AC020853 Mus muscu
C 863	17	1.2	4622	8	AF288216	AF288216 Oryza sat	936	17	1.2	19349	2	AC012969	AC012969 Drosophila
C 864	17	1.2	4646	14	MSSMVRNA	X73883 Wheat spind	937	17	1.2	20343	5	GGVTTTIG	GGVTTTIG Chiken vit
C 865	17	1.2	4726	3	AF210316	AF210316 Drosophila	938	17	1.2	21110	8	SPCC2386	SPCC2386 S.pombe c
C 866	17	1.2	4755	9	AB040914	AB040914 Homo sapi	939	17	1.2	23250	9	AC093016	AC093016 Homo sapi
C 867	17	1.2	4757	6	ARI45629	ARI45629 Sequence	940	17	1.2	23389	6	HSV526F1A	HSV526F1A Human DNA
C 868	17	1.2	4757	6	ARI45629	ARI45629 Sequence	941	17	1.2	23389	6	HSV526F1A	HSV526F1A Human DNA
C 869	17	1.2	5089	6	LES6378	U006378 Lycopersi	942	17	1.2	23489	9	AP001973	AP001973 Homo sapi
C 870	17	1.2	5156	6	AR085703	AR085703 Sequence	943	17	1.2	24833	9	AP001973	AP001973 Homo sapi
C 871	17	1.2	5401	10	MUSCDEPROA	M35525 Mouse compl	944	17	1.2	24903	3	CETLAC7	CETLAC7 Caenorhabd
C 872	17	1.2	5403	10	MUSCDEPROA	M35525 Mouse compl	945	17	1.2	26691	3	CETLAC7	CETLAC7 Caenorhabd
C 873	17	1.2	5542	9	HSU09954	U09954 Human ribos	946	17	1.2	27277	3	CELK0563	CELK0563 Caenorhabd

C 947	17	1.2	27332	3	CER11X8	Z73310 Caenorhabdi
948	17	1.2	27719	2	AC012623	AC012623 Homo sapi
949	17	1.2	27935	3	CELC07D10	-acetylglucosaminyltransferase e, c2gnf3
950	17	1.2	28409	3	U41012	Patent: WO 0114535-A 1 01-MAR-2001;
C 951	17	1.2	29688	3	CEC09G5	Schwentek, Tilo (DK)
952	17	1.2	29729	3	CELC13B9	Location/Qualifiers
953	17	1.2	30000	10	RNIGR2	1. 1362
C 954	17	1.2	30389	8	SPAC19D5	/organism="Homo sapiens"
955	17	1.2	30562	8	SPAC3A11	/db_xref="taxon:9606"
956	17	1.2	30885	1	AE001580	1. 1362
957	17	1.2	31857	2	AC020299	/note="The gene of C2GNF3"
958	17	1.2	33715	2	AC020075	/codon_start=-1
959	17	1.2	34110	2	AC020075	/protein_id="CAC34674.1"
960	17	1.2	34327	2	AC010426	/db_xref="GI:13396914"
961	17	1.2	34349	3	CELC10C2	/translation="MKIFKVFKEFTLOOKVFILFMTLSLKILNVRRLPRKDIY
962	17	1.2	34796	3	CEFS3C11	LVRELSSTSPFVNRRTTHYDEIRYVNGSITGEOPLEKSGLETRRDITDEND
963	17	1.2	34874	2	AC015453	VNMTSDCDITOTLRGKQQLVSKSEKSPFIATSLVHAKDAIVELHAIYKHNIT
964	17	1.2	35031	3	AF052833	CIHYDRAPDTEFVAMNNKLCFSNIFLSKLEAVEYAHISRLQADINCLSDLKSI
C 965	17	1.2	35409	3	CEC44C10	QWYVYIMLCGODEPLKSNFELVSEKLNANMLLEVYKPNKSLERFTYHEELRVY
966	17	1.2	35484	3	CEP26E4	EYVLPPIRINISKEAPPHNIQIFVGSAPFYKIFNNISVODFPWMSKDYSP
967	17	1.2	35579	9	H5359F1	DEHFWALIRVPGIPGHSRQDVSDLOSKRTLVKMYVEGFYSCGSHLSVCI
C 968	17	1.2	35762	8	AB026553	YGAALMLMLKDGHWPFANKRDSVDPDILIKCLAEKLEBQQRDMITLPSKLEFDRMLT
969	17	1.2	36384	9	AC002209	TTS"
970	17	1.2	36393	9	D87013	BASE COUNT 410 a 260 c 279 g 413 t
C 971	17	1.2	36972	9	AF198491	ORIGIN
C 972	17	1.2	37010	9	CEBG28M15	Query Match 100.0%; Score 1362; DB 6; Length 1362;
C 973	17	1.2	37170	9	HSE129H9	Best Local Similarity 100.0%; Pred. No. 0;
974	17	1.2	37302	9	D87007	Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
975	17	1.2	37329	3	U97009	
C 976	17	1.2	37329	3	U97009	
977	17	1.2	37515	3	CEBG47P01	
978	17	1.2	37628	3	CEP31D4	
979	17	1.2	37836	3	AC024695	
980	17	1.2	37836	3	AF045639	
C 981	17	1.2	38338	6	AX059493	
982	17	1.2	38433	3	U40942	
983	17	1.2	38553	3	CEH25K10	
C 984	17	1.2	38715	9	AC004237	
C 985	17	1.2	38950	3	CEP19B2	
C 986	17	1.2	39215	3	CEB1054	
C 987	17	1.2	39568	2	AC014823	
C 988	17	1.2	39702	7	AC002111	
989	17	1.2	39732	7	AF069529	
C 990	17	1.2	39781	7	SPC11739	
991	17	1.2	39910	9	AP001039	
C 992	17	1.2	39944	9	AC011541	
C 993	17	1.2	40014	3	CELF48B9	
994	17	1.2	40014	3	DMC87B1	
C 995	17	1.2	40301	3	U00065	
C 996	17	1.2	40466	3	CELB0496	
C 997	17	1.2	40633	2	AC017139	
998	17	1.2	41170	9	AP000951	
999	17	1.2	41306	9	AP000523	
1000	17	1.2	41587	3	AF052832	

ALIGNMENTS

RESULT 1						
LOCUS AX087935	1362 bp	DNA		17-MAR-2001		
DEFINITION Sequence 1 from Patent WO0114535.						
ACCESSION AX087935						
VERSION AX087935.1	GI:13396913					
KEYWORDS						
SOURCE human.						
ORGANISM Homo sapiens						
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
1 (bases 1 to 1362).						
1 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE Schwientek, T. and Clausen, H.						
1 udg-n-acetylglucosamine: galactose-g(b)1,3- n						
AUTHORS TITLE						

JOURNAL	-acetylglucosamine-g(a)-x/(g(a) to galnac)-g(b)1,6- n
FEATURES	-acetylglucosaminyltransferase e, c2gnf3
SOURCE	Patent: WO 0114535-A 1 01-MAR-2001;
	Schwentek, Tilo (DK)
	Location/Qualifiers
	1. 1362
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	1. 1362
	/note="The gene of C2GNF3"
	/codon_start=-1
	/protein_id="CAC34674.1"
	/db_xref="GI:13396914"
	/translation="MKIFKVFKEFTLOOKVFILFMTLSLKILNVRRLPRKDIY
	LVRELSSTSPFVNRRTTHYDEIRYVNGSITGEOPLEKSGLETRRDITDEND
	VNMTSDCDITOTLRGKQQLVSKSEKSPFIATSLVHAKDAIVELHAIYKHNIT
	CIHYDRAPDTEFVAMNNKLCFSNIFLSKLEAVEYAHISRLQADINCLSDLKSI
	QWYVYIMLCGODEPLKSNFELVSEKLNANMLLEVYKPNKSLERFTYHEELRVY
	EYVLPPIRINISKEAPPHNIQIFVGSAPFYKIFNNISVODFPWMSKDYSP
	DEHFWALIRVPGIPGHSRQDVSDLOSKRTLVKMYVEGFYSCGSHLSVCI
	YGAALMLMLKDGHWPFANKRDSVDPDILIKCLAEKLEBQQRDMITLPSKLEFDRMLT
	TTS"
BASE COUNT	410 a 260 c 279 g 413 t
ORIGIN	
Query Match	100.0%; Score 1362; DB 6; Length 1362;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1362;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1	atgaagatccaagtgtatattttaaacaataccctaacagagaagtgttcacgtctt 60
1	atgaagatccaagtgtatattttaaacaataccctaacagagaagtgttcacgtctt 60
61	tttaaccatggcgcgcctctcttctttaaagcttctaagtgtgagagacctcttcgcaa 120
61	tttaaccatggcgcgcctctcttctttaaagcttctaagtgtgagagacctcttcgcaa 120
121	aaagacattacttggttgaagtaactccctaagtaactgccttcttgaagaacaagatac 180
121	aaagacattacttggttgaagtaactccctaagtaactgccttcttgaagaacaagatac 180
181	actcgttgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 240
181	actcgttgaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 240
241	ccttggaaattggaagaagtcctggaataaagaagaagaagcattgagctgagat 300
241	ccttggaaattggaagaagtcctggaataaagaagaagaagcattgagctgagat 300
301	gattgattgttggaataagcagatgattggaatttgaagtaagtaagtaagtaagtaag 360
301	gattgattgttggaataagcagatgattggaatttgaagtaagtaagtaagtaagtaag 360
361	caaaagcttgctcaaaagagagagaagaagcttcccaataagcctatcttggttgcac 420
361	caaaagcttgctcaaaagagagagaagaagcttcccaataagcctatcttggttgcac 420
421	aaagatgaattatggttgaagaagcttaccatgctataatacaacagacaaatattac 480
421	aaagatgaattatggttgaagaagcttaccatgctataatacaacagacaaatattac 480
481	tgatccattatgatacgtaagagcaccctgataccttcaaaagttgcacatgaattagct 540
481	tgatccattatgatacgtaagagcaccctgataccttcaaaagttgcacatgaattagct 540
541	aagtgtcttccaatatttcttcttccaaataagagctgtggaatatgacacatt 600
541	aagtgtcttccaatatttcttcttccaaataagagctgtggaatatgacacatt 600
601	tcacagctcagagctgatttaaatgtctgtgcagaccttgcgaagcttcaatcagctg 660
601	tcacagctcagagctgatttaaatgtctgtgcagaccttgcgaagcttcaatcagctg 660

QY 661 aaatgttatacaactgtgttggaagaatttccocctgaagtcacaatttgaattggtg 720
 |||||
 Db 661 AAATATGTATCACTGTTGTGGGCAAGATTTCCCTCAAGTCAAAATTTGAATTGGTG 720
 QY 721 tcaaggttgaaaaaactcaatgtggaacaata-ctgttggaagacgtgttaaaccccaaacgt 780
 |||||
 Db 721 TCAGAGTTAAAAAACTCAATGAGCAAAATATGTTGGAAACGCTGAACCCCAACACAGT 780
 QY 781 aaattgaaagattcactaccatcatcagtaagacggtgtccttaagatgtgaa 840
 |||||
 Db 781 AAATTGAAAAGATTCTCTACCATCATGAACTTAGAGGGGCTCCCTATGAAATATGTGAAG 840
 QY 841 ctaccataagagaacaacatctcccaaggaagaccccccaataacattcagatttgtt 900
 |||||
 Db 841 CTAACCAATAGAGCAAAACATCTCCAAGGAAGACACCCCAATCAATTCAGATATTGTGT 900
 QY 901 ggcaggtctatttctttaaagtcagacattgtttaataataatttcaacaacccatc 960
 |||||
 Db 901 GGCAGGCTTATTCTTTTAAGTCAAGCATTTGTTAAATATTTTCAACAACCTCCATC 960
 QY 961 gtccaagacttttctgctgtgtcctaaagacatactctcctgataagacatttggct 1020
 |||||
 Db 961 GTTCAAGACTTTTTCCTGCTGCTAAAGACACATCTCTCTGATGAGCACTTTTGGGCT 1020
 QY 1021 acccttattccgggttccagaataactctggggaagatttccagaatcagccagatgtgt 1080
 |||||
 Db 1021 ACCTTATTCGGGTTCCAGGAATACCTGGGAGATTCCAGATCAAGCCAGATGTGTCT 1080
 QY 1081 gatctgcagagtaagaactgcgcttgcaggtgaattactaatagaagcttttctatccc 1140
 |||||
 Db 1081 GATCTGCAGAGTAAGACTGGCCTTGTCAAGTGAATTAAGTAAGAGGCTTTTCTATCC 1140
 QY 1141 agtctgacggagactcaactcctcgaagcgtgtgatttaagagcctcagaataagatg 1200
 |||||
 Db 1141 AGTTGACTGGATCTCACTTCGAACCGTGTGATTTAAGAGCTCAAAATTAAGGTG 1200
 QY 1201 ctatacaagatgacattgtgttcttaataaattgattcctaaggttgagccatctgt 1260
 |||||
 Db 1201 CTATATCAAGATGAGCATTTGGTTGCTAATTAATTTGATTTCAAGGTGAGACCTATCTTG 1260
 QY 1261 attaatgtctggcagaagaagcttgaagaacagcagaagacgtgatacttggcctca 1320
 |||||
 Db 1261 ATTAATATGCTTGGCAGAAAAGCTTTGAAGACAGACAGAGACAGATCTTGCTTGCCTCA 1320
 QY 1321 gaaaagttaatttctgataagaatactcactaccacatactga 1362
 |||||
 Db 1321 GAAAGTATTATTATGATTAGAAATCTACACTACACATCATGA 1362
 RESULT 2
 AF132035 3435 bp mRNA PRI 08-APR-2000
 LOCUS Homo sapiens core 2 beta-1,6-N-acetylglucosaminyltransferase 3
 DEFINITION (C2GNT3) mRNA, complete cds.
 ACCESSION AF132035
 VERSION AF132035.1 GI:7527463
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 3435)
 AUTHORS Schwientek, T., Yeh, J. C., Levery, S. B., Keck, B., Merix, G., van
 Kessel, A. G., Fukuda, M. and Clausen, H.
 TITLE Control of O-glycan branch formation. Molecular cloning and
 characterization of a novel thymus-associated core 2 beta-1,
 6-n-acetylglucosaminyltransferase
 J. Biol. Chem. 275 (15), 11106-11113 (2000)
 JOURNAL MEDLINE
 20219156
 REFERENCE 2 (bases 1 to 3435)
 AUTHORS Schwientek, T. and Clausen, H.
 TITLE Direct Submission

JOURNAL Submitted (27-FEB-1999) School of Dentistry, University of
 Copenhagen, Nørre Alle 20, Copenhagen 2200, Denmark
 FEATURES Location/Qualifiers
 source 1..3435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene 1..3435
 /gene="C2GNT3"
 862..2223
 /gene="C2GNT3"
 /codon_start=1
 /product="core 2 beta-1,6-N-acetylglucosaminyltransferase
 3"
 /protein_id="AF63156.1"
 /db_xref="GI:7527464"
 /translation="MKTFKCYFHHHTLOQKVFILFLIMLISLILIANVRLPRKNDY
 LVEYSSTSPFVNRNRYTHVADVEYRIVNCGSGLT EQLPLEIGKLELRRIIDLEDD
 VVAMTSQCDLYQTLRGYAOQLVKEKSESPILAVSLVHDAIWEHLIAHINQNIY
 CIHYDRKAPDPTFVAMNNLAKCFSNIFLASKLEAVYAHISLQADLNLSDLKSSI
 QMKVIVLGGDPEPLKSNFELVSELKLNKAMLEKVPNSKLEFTHHELRVPY
 EYKLPTRINISKEDAPHNIQIFVSAVYFLSQAPVKYIFNNSIYODFPAMSKOTSP
 DEHPMFLIRVPGTIPGELISRSADVDSDLSKRTLVKMYEGRFYSCTGSHLSVCI
 YGAELRWLTKDQHWFRANKFDSYVDPLILICLAKLEEDQRMWITLPSKLPDRMLT
 TTS"
 BASE COUNT 1032 a 678 c 651 g 1074 t
 ORIGIN
 Query Match 100.0%; Score 1362; DB 9; Length 3435;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgaagatattcaaatgttattttaaacaataccctacagcagaagtttcatcctgtt 60
 Db 862 ATGAAGATATTCAAATGTTATTTTAAACATACCTACAGCAGAAAGTTTCATCTGTTT 921
 QY 61 ttaacccatgctgctctcttctttaaagcttctaagtgaagcagctcttcggcaa 120
 Db 922 TTAACCCATAGGCTGCTCTCTTGTAAAGCTCTCAATATGAGACGACTTTTCCGCA 981
 QY 121 aaagacattacttctgtgagtaactccctaaagtaactgccttcttgaagaacaga 180
 Db 982 AAAGACATTTACTTGTTGAGTACTCCCTAAGTACCTGCGCTTTGTGAAGAAACAGATAC 1041
 QY 181 actcatgttaagatgaagatgaagcagatgaagtaactcttgggtatcataagaacag 240
 Db 1042 ACTCATGTTAAAGATGAAGTCAAGTATACGTGTGGGATCTATGAACAGGAG 1101
 QY 241 ccttggaaattggaagaagctctggaataagaagaaggaacatcatgacttggagat 300
 Db 1102 CCTTGGAAATTGGAAGAAGTCTGGAATTAAGAAAGGACATCATTGACTTGAGAGAT 1161
 QY 301 gatgatgttggaagaagcagatgaatgtgacatttcaagactctaagaagttatgct 360
 Db 1162 GATGATGTTGGCAATGCCAGTATGTGCAATTATCAAGACTCTAAGAGGTTATGCT 1221
 QY 361 caaagcgtgtctcaagaagagagaagaagcttcccaatagccatcttctgtgtccac 420
 Db 1222 CAAAAGCTTGTCTCAAAGAGAGAAAAGCTTCCCAATAGCCTATTCCTTGTTGTCAC 1281
 QY 421 aaagatgcaataatgttggaagcgtatccatgctatatacaacagcacacaattaac 480
 Db 1282 AAAGATGCAATATATGTTGAAGAGGCTTATCCATGTATATACAAACACACAAATTTTAC 1341
 QY 481 tgcattcatatgacgtgaagcagctatatactcctcaagttgcatgaaacaattagct 540
 Db 1342 TGCATTCATTAATGATGTAAGCAGCTGATACCTTCAAAGTTGCCATGAACAATTACT 1401
 QY 541 aagtgcttccaataatttcatcttccaataatgaagcgtgtggaatatgcccacatt 600
 Db 1402 AAGTCTCTCCAAATATTTTCAATGCTTCCAAATATGAGAGGCTGTGAATATGCCACATT 1461
 QY 601 tccagctccagcgtgatttaaatgtctgctgacctctgaagttcttcaatcagctg 660

```

Db 1462 TCACAGCTCCAGCTGATTAAATGCTTGCGAGCTTCTAACTTCATCCAGTGG 1521
|||
Qy 661 aatatgttcaactgtgtggaagaatttccctgaatcaatttgaatgtg 720
|||
Db 1522 AATATNGTATCACTGTGTGGGCAAGATTTCCCTGAAGTCAATTTGAAATGGTG 1581
|||
Qy 721 tcagagtgaaaaaactcaatggagaatatgttggagaagggtgaaccccaacagt 780
|||
Db 1582 TCAGAGTTGAAAAAACTCAATGAGCAAAATATGTGGAGACGCTGAACCCCAACAGT 1641
|||
Qy 781 aaattgaaagattccttaacatcagactaagacgggtgcttaataatgtgaag 840
|||
Db 1642 AATTTGAAAGATTCCTTACCTCTCATGTAGATTAGACGGTCCCTTAATGAAATGTGAAG 1701
|||
Qy 841 ctaccaaagaagaacaactcctcaagaagacccccccataaagattcagatttgt 900
|||
Db 1702 CTACCAATAGAGCAACATCTCCAAAGAGCACCCCCCATACATTCAGATATTGTGT 1761
|||
Qy 901 ggaagtgctattttgttaagtaagacattgttaaatatatttcaaacacacac 960
|||
Db 1762 GGCAGTGTATTATTTGTAAAGTCAAGCATTTGTTAAATATTTTCAACAACTCCATC 1821
|||
Qy 961 gtcaagacttttctgctcgtgtcaagaacacatactcctcgtatgagcactttgagct 1020
|||
Db 1822 GTTCAAGACTTTTCTGCTGCTAAAGACACATCTCTCGATGAGCAGCTTTGGGCT 1881
|||
Qy 1021 acctgattcgggttccgaagaactggggagatttccagttccagccagatgtgtc 1080
|||
Db 1882 ACCTTATTCGGGTTCAGGAATACTGGGGAGATTTCCAGATCACCCAGATGTGTGT 1941
|||
Qy 1081 gactcagaagtaagactcgccttgcgaagtgaattactaagaagcttttctatccc 1140
|||
Db 1942 GATCTGAGAGTAAGACTGCCCTTGTCAAGTGAATTACATGAAGCCTTTTCTATCCC 2001
|||
Qy 1141 agttgactgagatcactccttgaagcgtgtgtattttagggagctcaaatagaagtgt 1200
|||
Db 2002 AGTTGATCTGATCTCACTTCAGACGCTGTGATTATGAGACTGCACAAATTAAGTGTG 2061
|||
Qy 1201 cttaacaaagatgagacattgttcttaataattgacttaagttgagaccatctgt 1260
|||
Db 2062 CTATATCAAGATGAGACTTGTGTTGCTAATTAATTGATTCATAGGTGACCTATCTGTG 2121
|||
Qy 1261 ataaatgcttgagcaagaagcttgaagaagacagagagactgagctacttggccctca 1320
|||
Db 2122 ATTAATAAGCTTGGCAGAAAGCTTGAAGAACAGCAGAGACATGATCACTTGGCCCTCA 2181
|||
Qy 1321 gaaaagttattatgtagatagaatcactacacacatcatga 1362
|||
Db 2182 GAAAGTTATTATGATAGATAAATCTCACACACATCATGA 2223
|||

```

```

RESULT 3
AC093259/c DNA HTG 16-AUG-2001
LOCUS Homo sapiens chromosome 5 clone RP11-229C3, WORKING DRAFT SEQUENCE,
DEFINITION 6 unordered pieces.
AC093259.1 GI:15193393
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 184590)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 184590)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

```

COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 494352
Center clone name: RP11-229C3
-----
Summary Statistics
Consensus quality: 176980 bases at least Q40
Consensus quality: 182494 bases at least Q30
Consensus quality: 182978 bases at least Q20
Estimated insert size: 183600; agarose-ef estimation
Estimated insert size: 184090; sum-of-contigs estimation
Quality coverage: 8.93 in Q20 bases; agarose-ef estimation
Quality coverage: 8.9 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2136: contig of 2136 bp in length
* 2137 2236: gap of unknown length
* 2237 5820: contig of 3584 bp in length
* 5821 5920: gap of unknown length
* 5921 8425: contig of 2505 bp in length
* 8426 8525: gap of unknown length
* 8526 42924: contig of 34399 bp in length
* 42925 43024: gap of unknown length
* 43025 98717: contig of 55693 bp in length
* 98718 98817: gap of unknown length
* 98818 184590: contig of 85773 bp in length.
*
* Location/Qualifiers
* source
*   1..184590
*     /organism="Homo sapiens"
*     /db_xref="taxon:9606"
*     /chromosome="5"
*     /clone="RP11-229C3"
*     /clone_1fb="RP11 human BAC library 11"
BASE COUNT 56307 a 36971 c 36814 g 53966 t 502 others
ORIGIN
Query Match 92.5%; Score 1260; DB 2; Length 184590;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 22172 GATGATGTTGGCAATGACACCGATGATTGTGACATTTATCAGACTCTAAGAGGTTTNGCT 22113
Qy 361 caaaagctgtctcaaaagagagagaaagcttcccaatagcctattcttggttccac 420
Db 22112 CAAAAGCTTGTCTCAAAGAGAGAGAAAAGCTTCCCAATAGCCTATTCTTGGTTGTCCAC 22053
Qy 421 aaagatgcaatlatggttgaagagctlatccatgtctatatataacacagacacataattac 480
Db 22052 AAAGATGCAATATGTTGTAAGAGGCTTATCCATGCTATATACACACACACAAATATTATAC 21993
Qy 481 tgcacacatlatgacgtgaagacacgtatattacattcaaaagttgcacatgaacatttact 540
Db 21992 TGCATCATATTATGATCGTAGAGACCTGATACCTTCCTCAAGATTGCCATGAACATTTACT 21993
Qy 541 aagtgcttcccaatatttccattgtcttccaaattagagagctgtgtgaatgtgccacatt 600
Db 21932 AAGTGCTCTCCAAATTTTTCATGCTTCCAAATTAGAGGCTGTGGAATATGCCACATTT 21873
Qy 601 tccagatccagagctgatttaattgctgtgtgagaccttgcgaacttcccaatccagtg 660
Db 21872 TCCAGACTCCAGGCTGATTTAAATGCTGTGTGACCTTCTGAAGCTTCCATCCAGTGG 21813
Qy 661 aaatatgtatcaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 720
Db 21812 AAATATGTTATCACTTGTGTGGGCAAGATTTCCCTCAAGTCAAAATTTGAATTTGGTG 21753
Qy 721 tccagatgtgaaaaaactcaatgtgagcaaatatgtgtgtgtgtgtgtgtgtgtgtgtgt 780
Db 21752 TCCAGATTTAAAAAATCTCATGTGAGCAATATGTTGAGACCGTGTGAACCCCAACAGT 21693
Qy 781 aaatgtgaaagatcttaccatcatcatcatcatcatcatcatcatcatcatcatcatcat 840
Db 21692 AAATGTGAAGAATTCATCTTACCATCATGAACTTAGACGGGTGCCATTATGATGTGAAG 21633
Qy 841 ctaccataagagacaacatctccaaagagagacaccccccaataacattagatatttgt 900
Db 21632 CTACCAATAGAGCAAAACATCTCCAAAGAGAGACACCCCAATACATTCAGATTTTGT 21573
Qy 901 ggcagtgctatttctttaaagcaagcatgtgttaatatatttcaacaactccatc 960
Db 21572 GGCAGTGCTATTATTGTTTAAAGTCAAGATTGTTAAATATTTCACAACATCCATC 21513
Qy 961 gtccaagacttttgcctgtgtctaaagacacatctctcgtgagagcatttggagct 1020
Db 21512 GTTCAAGACTTTTTCCTCGGGTCTAAAGACATATCTCCGATGAGCAGCTTTTGGGCT 21453
Qy 1021 acctgattcgggttccaaagatcctgtggagagatttccagatccagccagagatgtct 1080
Db 21452 ACCATTATTCGGGTTCCAGGAATCTGGGGAGATTTCAGATCCAGCCAGATGTGT 21393
Qy 1081 gactgcagagtaagactgcgctgtcaagtggaattactatgaaggttttctatccc 1140
Db 21392 GATCTCAGAGTAAGACTGCGCTTGTCAAGTGAATTAATGAAAGGCTTTTCTATCC 21333
Qy 1141 agtgtgactggatctcaccttcaagcgtgtgtatttgaagagctcagaattaaagtgtg 1200
Db 21332 AGTGTGACTGGATCTCAACCTTCGAGCGTGTGATTTATGAGACCTCCAGAAATTAAGTGG 21273
Qy 1201 ctatacaagatggaattgtgtgtctataaattgattcttaagggtggaacctatctg 1260
Db 21272 CTATCAAAAGATGACATTTGTTGCTAATTAATTTGATTTAGGTGAGACCTATCTTG 21213
Qy 1261 attaatgtcttgcaaaaagcttgaagacagcagaagagctgggttccttgcctcta 1320
Db 21212 ATTAAATGCTTGCAAAAAGCTTGAAGACACAGACAGAGACTGGATCACTTGGCCCTCA 21153
Qy 1321 gaaaagtatttcttgatagaatactcactacacacatcatga 1362
Db 21152 GAAAAGTTATTATGATAGAAATTCACATACACACATCATGA 21111

```

RESULT 4
AX087939

```

LOCUS AX087939 31 bp DNA PAT 17-MAR-2001
DEFINITION Sequence 5 from Patent WO0114535.
ACCESSION AX087939
VERSION AX087939.1 GI:13396917
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 31)
AUTHORS Schwiientek, T. and Clausen, H.
TITLE Udp-n-acetylglucosamine-6-(a)-r/(g)lucac to galnac -g(b)1,6- n
JOURNAL -acetylglucosaminyltransferase e. c2gnr3
PATENT: WO 0114535-A 5 01-MAR-2001.
Schwiientek, Tilo (DK) ; Clausen, Henrik (DK)
FEATURES
source
1. .31
Location/Qualifiers
BASE COUNT 10 a 6 c 7 g 8 t
ORIGIN
Query Match 1.8%; Score 25; DB 6; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 114 tccgcaaaagacatttactgtgt 138
Db 7 TCCGCAAAAGACATTTACTTGCTT 31

```

RESULT 5
AX087937/c 21 bp DNA PAT 17-MAR-2001
LOCUS AX087937
DEFINITION Sequence 3 from Patent WO0114535.
ACCESSION AX087937
VERSION AX087937.1 GI:13396915
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 21)
AUTHORS Schwiientek, T. and Clausen, H.
TITLE Udp-n-acetylglucosamine: galactose-g(b)1,3- n
JOURNAL -acetylglucosamine-g(a)-r/(glucac to galnac) -g(b)1,6- n
PATENT: WO 0114535-A 3 01-MAR-2001.
Schwiientek, Tilo (DK) ; Clausen, Henrik (DK)
FEATURES
source
1. .21
Location/Qualifiers
BASE COUNT 5 a 7 c 3 g 6 t
ORIGIN
Query Match 1.5%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
CNSOLAY6 696 bp mRNA PLN 02-SEP-1999
LOCUS CNSOLAY6
DEFINITION Botrytis cinerea strain T4 cDNA library under conditions of
nitrogen deprivation.

ACCESSION AL113542.1 GI:5828161
 VERSION CDNA library; nitrogen deprivation.
 KEYWORDS Botryotinia fuckeliana.
 SOURCE Botryotinia fuckeliana
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes; Helotiales; Sclerotiniaceae; Botryotinia.
 REFERENCE 1 (bases 1 to 696)
 AUTHORS Bilton, F., Lewis, C., Fortini, D., Pradier, J.M. and Brygoo, Y.
 TITLE Direct Submission
 JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
 REFERENCE 2 (bases 1 to 696)
 AUTHORS Genoscope.
 LOCUS AC023731/c
 DEFINITION Direct Submission
 TITLE Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage; CP 5706 91057 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 COMMENT The CDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.
 FEATURES
 source 1. 696
 /organism="Botryotinia fuckeliana"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID: W51C051"
 BASE COUNT 186 a 158 c 164 g 188 t
 ORIGIN
 Query Match 1.5%; Score 21; DB 8; Length 696;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 243 ttgtgaattggaaagagctct 263
 ||||||||||||||||||
 Db 587 ttgtgaattggaaagagctct 607
 RESULT 7
 AC014188 9643 bp DNA HTG 16-NOV-1999
 LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AC014188
 VERSION AC014188.1 GI:6437147
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 9643)
 AUTHORS Adams, M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10212574 by the submitter. For further information on this sequence e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced.
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source 1. 9643
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 3212 a 1955 c 1666 g 2810 t
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 9643;
 Best Local Similarity 100.0%; Pred. No. 4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 520 gtctccatgacattagct 540
 ||||||||||||||||||
 Db 5082 gtctccatgacattagct 5102
 RESULT 8
 AC023731/c
 LOCUS Drosophila melanogaster clone RP98-33N24, *** SEQUENCING IN PROGRESS ***, 20 unordered pieces.
 DEFINITION
 ACCESSION AC023731.2 GI:6996819
 VERSION AC023731
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 24650)
 AUTHORS Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bodola, B., Bouck, J., Bowie, S., Brooks, A., Bulay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guetara, M., Harris, K., Hernandez, J., Hodgson, A., Jia, Y., Jones, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Owsal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L., Quesal, M., Reiter, D., Rives, M., Sammel, S., Say, J., Scherer, S., Shan, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Suong, R., Taber, P., Taylor, T., Vasquez, J., Vilson, R., Vo, Q., Wabshan, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 24650)
 TITLE Drosophila melanogaster
 REFERENCE
 AUTHORS
 JOURNAL
 COMMENT
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: DRJ
 Center clone name: RP98-33N24
 Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 98% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 7321 bases at least Q40
 Consensus quality: 13196 bases at least Q30
 Consensus quality: 16806 bases at least Q20
 Estimated insert size: 19923; sum-of-contigs estimation
 Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 20 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 834: contig of 834 bp in length
835 854: gap of unknown length
855 1694: contig of 840 bp in length
1695 1714: gap of unknown length
1715 2700: contig of 986 bp in length
2701 2720: gap of unknown length
2721 3985: contig of 1265 bp in length
3986 4005: gap of unknown length
4006 5224: contig of 1219 bp in length
5225 5244: gap of unknown length
5245 6444: contig of 1200 bp in length
6445 7897: gap of unknown length
7898 7917: gap of unknown length
7918 9148: contig of 1831 bp in length
9149 9168: gap of unknown length
9169 10197: contig of 1029 bp in length
10198 10217: gap of unknown length
10218 11717: contig of 1500 bp in length
11718 11737: gap of unknown length
11738 12664: contig of 927 bp in length
12665 12684: gap of unknown length
12685 13429: contig of 745 bp in length
13430 13449: gap of unknown length
13450 14854: contig of 1405 bp in length
14855 14874: gap of unknown length
14875 16285: contig of 1411 bp in length
16286 16305: gap of unknown length
16306 17671: contig of 1366 bp in length
17672 18998: contig of 1307 bp in length
18999 19018: gap of unknown length
19019 20156: contig of 1138 bp in length
20157 20176: gap of unknown length
20177 21638: contig of 1462 bp in length
21639 21658: gap of unknown length
21659 22795: contig of 1137 bp in length
22796 22815: gap of unknown length
22816 24650: contig of 1835 bp in length.
  
```

FEATURES
 source
 1. 24650
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RP98-33N24"

BASE COUNT 6800 a 4962 c 5240 g 7211 t 437 others
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 24650;
 Best Local Similarity 100.0%; Pred. No. 4.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 520 gttgcacatgaacaattagct 540
 ||||||||||||||||||
 Db 20732 GTTGCATGACACATTTAGCT 20712

RESULT 9
 AP002084 84011 bp DNA HTG 25-JUL-2000
 LOCUS Homo sapiens chromosome 4 clone 2179J-8 map 4q22-q24, *** SEQUENCING
 IN PROGRESS ***; 13 unordered pieces.
 ACCESSION AP002084
 VERSION AP002084.2 GI:9453892
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Homo sapiens DNA, clone:2179L8.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (19-MAY-2000) Shih-Feng Tsai, National Yang-Ming
 University, Institute of Genetics, 155 Li-Kong St. Section 2,
 Beitou, Taipei, Taiwan 11221, Republic of China
 (E-mail: ympeisai@ym.edu.tw, URL: http://genome.ym.edu.tw/
 Tel: 886-2-28267043, Fax: 886-2-28264930)
 On Jul 25, 2000 this sequence version replaced gi:7959372.
 These sequences are draft human sequences, not finished sequences.
 Of the same clone are represented as 100 N.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

COMMENT

```

1 833: contig of 833 bp in length
834 934: gap of 100 bp
934 934: 1718: contig of 785 bp in length
1719 1818: gap of 100 bp
1819 6473: contig of 4655 bp in length
6474 6573: gap of 100 bp
6574 12038: contig of 5465 bp in length
12039 12138: gap of 100 bp
12139 16985: contig of 4847 bp in length
16986 17085: gap of 100 bp
17086 22568: contig of 5483 bp in length
22569 22668: gap of 100 bp
22669 28193: contig of 5525 bp in length
28194 28293: gap of 100 bp
28294 35735: contig of 7442 bp in length
35736 35835: gap of 100 bp
35836 46541: contig of 10706 bp in length
46542 46641: gap of 100 bp
46642 53098: contig of 6449 bp in length
53099 53190: gap of 100 bp
53191 61583: contig of 8393 bp in length
61584 61683: gap of 100 bp
61684 73057: contig of 11374 bp in length
73058 73157: gap of 100 bp
73158 84011: contig of 10854 bp in length.
  
```

FEATURES
 source
 1. 84011
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4q22-q24"
 /clone="2179L8"

BASE COUNT 26194 a 14705 c 15392 g 26513 t 1207 others
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 84011;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1073 atgtctcgtatcgcagagta 1093
 ||||||||||||||||||
 Db 22215 ATGTCTCGATCTGCGAGGTA 22235

RESULT 10
 AC011375 133008 bp DNA HTG 05-MAY-2000
 LOCUS Homo sapiens chromosome 5 clone CTS-119L1, WORKING DRAFT SEQUENCE.
 DEFINITION 10 unordered pieces.
 ACCESSION AC011375

VERSION AC011375.4 GI:7710555
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 133008)
 AUTHORS DOE Joint Genome Institute.
 TITLES Sequencing of Human Chromosome 5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 133008)
 AUTHORS DOE Joint Genome Institute.
 TITLES Direct Submission
 JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On May 5, 2000 this sequence version replaced gi:6597091.
 COMMENT -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

Project Information
 Center Project Name: 110165, H275
 Center clone name: CIT978SKB_119L1

Summary Statistics
 Consensus quality: 121552 bases at least Q40
 Consensus quality: 127643 bases at least Q30
 Consensus quality: 129058 bases at least Q20
 Estimated insert size: 140000; pulse field gel estimation
 Estimated insert size: 132108; sum-of-contigs estimation
 Quality coverage: 6.28 in Q20 bases; pulse field gel estimation
 Quality coverage: 6.65 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2822: contig of 2822 bp in length
 * 2823 2922: gap of unknown length
 * 2923 5756: contig of 2834 bp in length
 * 5757 5856: gap of unknown length
 * 5857 9902: contig of 4046 bp in length
 * 9903 10002: gap of unknown length
 * 10003 12177: contig of 2175 bp in length
 * 12178 12277: gap of unknown length
 * 12278 19083: contig of 6806 bp in length
 * 19084 19183: gap of unknown length
 * 19184 29949: contig of 10766 bp in length
 * 29950 30049: gap of unknown length
 * 30050 47593: contig of 17544 bp in length
 * 47594 47693: gap of unknown length
 * 47694 70445: contig of 22752 bp in length
 * 70446 70545: gap of unknown length
 * 70546 95942: contig of 25397 bp in length
 * 95943 96043: gap of unknown length
 * 96043 133008: contig of 36966 bp in length.

FEATURES
 Location/Qualifiers
 1. 133008

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"

/clone="CTB-119L1"
 /clone_id="Caltech human BAC library B"

BASE COUNT 39430 a 25488 c 25495 g 41692 t 903 others
 ORIGIN

Query Match 1.5%; Score 21; DB 2; Length 133008;
 Best Local Similarity 100.0%; Pred. No. 5.1;

Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1243 aaggtgcacctatcttgatt 1263
 Db 20405 AAGGTGCACCTATCTTGATT 20425

RESULT 11

AC073826

LOCUS AC073826 146190 bp DNA HTG 29-JUN-2000
 DEFINITION Mus musculus clone RP23-93K16, WORKING DRAFT SEQUENCE, 35 unordered

ACCESSION

AC073826
 AC073826.1 GI:8810443
 HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS
 SOURCE
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS
 TITLES

REFERENCE
 AUTHORS
 TITLES
 JOURNAL

2 (bases 1 to 146190)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

Project Information
 Center Project Name: 1772840
 Center clone name: RPCI-23_93K16

Summary Statistics
 Consensus quality: 118979 bases at least Q40
 Consensus quality: 131063 bases at least Q30
 Consensus quality: 133821 bases at least Q20
 Estimated insert size: 187180; agarose-gel estimation
 Estimated insert size: 142790; sum-of-contigs estimation
 Quality coverage: 8.66 in Q20 bases; agarose-gel estimation
 Quality coverage: 11.35 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 * consists of 35 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1067: contig of 1067 bp in length
 * 1068 1167: gap of unknown length
 * 1168 2519: contig of 1352 bp in length
 * 2520 2619: gap of unknown length
 * 2620 3814: contig of 1195 bp in length
 * 3815 3914: gap of unknown length
 * 3915 5132: contig of 1218 bp in length
 * 5133 5232: gap of unknown length
 * 5233 6499: contig of 1267 bp in length
 * 6500 6599: gap of unknown length
 * 6600 7609: contig of 1010 bp in length
 * 7610 7710: gap of unknown length
 * 7710 8792: contig of 1083 bp in length
 * 8793 8892: gap of unknown length
 * 8893 10240: contig of 1348 bp in length
 * 10241 10340: gap of unknown length
 * 10341 11532: contig of 1192 bp in length
 * 11533 11633: gap of unknown length
 * 11633 13022: contig of 1390 bp in length
 * 13023 13122: gap of unknown length

```

ACCESSION      AL358777
VERSION        AL358777.12
KEYWORDS       HG.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Ekkaroyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Euthera; Primates; Catarrhini; Homiindrae; Homo.
TITLE          1 (bases 1 to 155359)
JOURNAL        Wallis,J.
COMMENT        Direct Submission
               Submitted (30-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
               C10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
               Request: clonerequest@sanger.ac.uk
               Oct 1, 2000 this sequence version replaced gi:10039983.
               During sequence assembly data is compared from overlapping clones.
               Where difference are found these are annotated as variations
               together with a note of the overlapping clone name. Note that the
               variation annotation may not be found in the sequence submission
               corresponding to the overlapping clone, as we submit sequences with
               only a small overlap as described above.
               This sequence has been finished according to sequence map criteria
               as follows. An attempt is made to resolve all sequencing problems,
               such as compressions and repeats, but not necessarily within known
               annotated human repeat sequence elements (e.g. Alu). Where the
               sequence is ambiguous, there is an annotation using the 'unsure'
               feature key.
               The following abbreviations are used to associate primary accession
               numbers given in the feature table with their source databases:
               Em:, EMBL, SW:, SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information
               Em: the WORMPEP database can be found at
               http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
               was generated from part of bacterial clone contigs of human
               chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
               Group. Further information can be found at
               http://www.sanger.ac.uk/HGP/chr6
               RP11-421M1 is from the library RPCR-11.2 constructed at the Roswell
               Park Cancer Institute by the group of Pletzer de Jong. For further
               details see http://bacpac.med.buffalo.edu/
               VECTON: pBac3.6
               IMPORTANT: This sequence is not the entire insert of clone
               RP11-421M1. It may be shorter because we sequence overlapping
               sections only once, except for a 100 base overlap.
               The true left end of clone RP11-421M1 is at 1 in this sequence. The
               true left end of clone RP3-417M14 is at 155260 in this sequence.
               The true right end of clone RP11-360019 is at 66388 in this
               sequence.
               Location/Qualifiers
                 1..155359
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="6"
                   /clone="RP11-421M1"
                   /clone_1bp="RPCR-11.2"
                   155289..155291
                   /note="clone RP3-417M14
                   gac in this entry
                   substitution
                   /replace="gcc"
BASE COUNT     42337 a 33067 c 33233 g 46732 t
ORIGIN
Query Match    1.5%; Score 21; DB 9; Length 155359;
Best Local Similarity 100.0%; Freq. NO. 5.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 597 cattccagactccagactga 617
| | | | | | | | | | | | | | |
Db 9925 cattccagactccagactga 9945

RESULT 13
RC023461

```


LOCUS AC023461 161676 bp DNA HTG 11-AUG-2000
 DEFINITION Homo sapiens chromosome 4 clone RP11-122J10 map 4, WORKING DRAFT
 AC023461
 AC023461.3 GI:9795645
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 161676)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 4, clone RP11-122J10
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 AUTHORS
 2 (bases 1 to 161676)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
 Dearielano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
 Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Coveite,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Landers,T., Laroque,K., Lehoczy,J., Levine,R.,
 Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
 McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
 Menes,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
 Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,S.,
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
 Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zairoun,J., Zimmer,A. and
 Zody M.
 Direct Submission
 Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 11, 2000 this sequence version replaced gi:7139847.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6815
 Center clone name: 122.J.10
 ----- Summary Statistics
 Sequencing vector: M13; M77815, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 154216 bases at least Q40
 Consensus quality: 157284 bases at least Q30
 Consensus quality: 159212 bases at least Q20
 Insert size: 150000; agarose-efp
 Insert size: 160576; sum-of-ctnigs
 Quality coverage: 8.0 in Q20 bases; agarose-efp
 Quality coverage: 7.5 in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order in this sequence record is
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 5459: contig of 5459 bp in length
 * 5460 5559: gap of 100 bp

FEATURES
 SOURCE
 * 5560 6680: contig of 1121 bp in length
 * 6681 6780: gap of 100 bp
 * 6781 7839: contig of 1059 bp in length
 * 7840 7939: gap of 100 bp
 * 7940 8959: contig of 1020 bp in length
 * 8960 9059: gap of 100 bp
 * 9060 10335: contig of 1276 bp in length
 * 10336 10435: gap of 100 bp
 * 10436 11687: contig of 1252 bp in length
 * 11688 11787: gap of 100 bp
 * 11788 13193: contig of 1406 bp in length
 * 13194 13293: gap of 100 bp
 * 13294 24793: contig of 11500 bp in length
 * 24794 24893: gap of 100 bp
 * 24894 47242: contig of 22349 bp in length
 * 47243 47342: gap of 100 bp
 * 47343 97906: contig of 50564 bp in length
 * 97907 98006: gap of 100 bp
 * 98007 136346: contig of 38340 bp in length
 * 136347 136446: gap of 100 bp
 * 136447 161676: contig of 25230 bp in length.
 Location/Qualifiers
 1. 161676
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-122J10"
 /clone_1lb="RPC1-11 Human Male BAC"
 1. 5459
 /note="assembly-fragment"
 /clone_end:596
 /vector_side:left"
 5560. 6680
 /note="assembly-fragment"
 6781. 7839
 /note="assembly-fragment"
 7940. 8959
 /note="assembly-fragment"
 9060. 10335
 /note="assembly-fragment"
 10436. 11687
 /note="assembly-fragment"
 11788. 13193
 /note="assembly-fragment"
 13294. 24793
 /note="assembly-fragment"
 24894. 47242
 /note="assembly-fragment"
 47343. 97906
 /note="assembly-fragment"
 98007. 136346
 /note="assembly-fragment"
 /note="assembly-fragment"
 /clone_end:77
 /vector_side:right"
 51145 a 27906 c 27160 g 54365 t 1100 others
 ORIGIN
 BASE COUNT 51145 a 27906 c 27160 g 54365 t 1100 others
 Query Match 1.5%; Score 21; DB 2; Length 161676;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1073 atgtctcatctgcagacta 1093
 ||||||||||||||||||||
 Db 27171 ATGTCTCATCTGCAGACTA 27191
 RESULT 14
 HSI68L15
 LOCUS HSI68L15 178098 bp DNA PRI 23-NOV-1999

DEFINITION Human DNA sequence from clone 168L15 on chromosome 6q26-27 Contains part of RPS6KA2 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3)), ESTs, STS, GSSs and Cpg islands, complete sequence.

ACCESSION AL022069

VERSION AL022069.1 GI:3256174

KEYWORDS HMG: Cpg Island, RPS6KA2.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 178098)

AUTHORS Palmer S.

TITLE Direct Submission

JOURNAL Submitted (24-JUN-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requesters: clonerequests@sanger.ac.uk

On Jun 26, 1998 this sequence version replaced gi:2950404.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

EM: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is the entire insert of clone 168L15. The true right end of clone 427A4 is at 13623 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

168L15 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Plerier de Jong. For further details see http://bacpac.med.buffalo.edu/VECTOR: pcvpac2.

FEATURES

source

1..178098

location/Qualifiers

1..178098

organism="Homo sapiens"

db_xref="taxon:9606"

chromosome="6"

map="q26-27"

clone_1lb="RPCI-1"

clone="RPI-168L15"

712..1186

repeat_region

/note="MER42 repeat: matches 1..504 of consensus"

1810..1812

variation

/note="clone 427A4; CCG in this entry; substitution"

/replace="cag"

2594..2661

repeat_region

/note="12 repeat: matches 2579..2660 of consensus"

3435..3437

variation

/note="clone 427A4; AAT in this entry; substitution"

/replace="aat"

3703..3916

repeat_region

/note="MER30 repeat: matches 1..230 of consensus"

3830..3832

variation

/note="clone 427A4; GCT in this entry; substitution"

/replace="gat"

3917..4181

misc_feature

/note="match: GSS: Em:B4242"

4418..5138

repeat_region

/note="LIMB3 repeat: matches 5461..6185 of consensus"

5594..5875

repeat_region

/note="LIMC4 repeat: matches 7120..7423 of consensus"

5917..6000

repeat_region

/note="LIMD3 repeat: matches 7345..7432 of consensus"

repeat_region

6124..6452

/note="LIMC4 repeat: matches 7671..7977 of consensus"

6348..6349

variation

/note="clone 427A4; TG in this entry; deletion"

/replace="tatataatg"

join(<6546..6625,8621..8700,12291..12397,15979..16016,18253..18395,26162..26232,27982..28070,46998..47062,57352..57454,65670..65800,68054..68179,84413..84502,86292..86450,93486..93647,96945..97021,98561..98678,102972..103109,104016..104561)

/gene="RPS6KA2"

/note="ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3); match: CDNAS: Em:L07597 Em:M2848 Em:M20188 Em:M20187 Em:M28489 Em:X85106 Em:U08316 Em:M99169 Em:L07598 Em:L07599 Em:D83013 Em:A0007938 Em:AB016869 Em:AB019245 Em:M58340; match: ESTs: Em:N79949 Em:F06080"

/evidence=not_experimental

/product="d168L15.1 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3))"

6546..104561

/gene="RPS6KA2"

join(<6546..6625,8621..8700,12291..12397,15979..16016,18253..18395,26162..26232,27982..28070,46998..47062,57352..57454,65670..65800,68054..68179,84413..84502,86292..86450,93486..93647,96945..97021,98561..98678,102972..103109,104016..104561)

/gene="RPS6KA2"

/note="match: proteins: SW:P18654 SW:P51812 SW:Q15349 Wp:CEH8169"

/codon_start=2

/evidence=not_experimental

/product="d168L15.1 (ribosomal protein S6 kinase, 90kD, polypeptide 2 (RSK3))"

/protein_id="CAB58228.1"

/db_xref="GI:6065881"

/translation="SGPSEIDEDGRH/GRESEPLHCEASISFOTEGKLYLIDFLRGG DLPRLSKENFTPEDEPKYALALALDHISLIGIYIDPENTLDEBHIKTD FGSKREAIHDHDKRATSCGTIEYMADEVYVRNRHOSADMSGVMPEMLTGSIDPF GVKDEETMALILKAKJLMPOLSLGEAOSILRALFKRNPCKRDAGIDGYEIKRHFF VTDWNTLYRKREIKPEPKPAVGRPEDETFHDPETLRHFDGEIKEDIGVSVCKRVH SFVAASLIQPSOODLHKVPVHPVIOQLHGNNTLHFDGEIKEDIGVSVCKRVH KADTEYAVAKIIDKSKRDPSEIEILRYOGHNTLTDKYVDGKFLVYLKMGK ELDIRILROYFESERESDVLCITITKMDYLHSGVYVHRDLKSNLILYDESGSPSI RLDQFGAKOLRAGNGLMPTCYTANVAEVIKRGQYDADCIISLIGILYTMLAGF TPRANGDDEPPEETIARISGKALSGMNDTSDAKDVSKMLVDDPHORTIAOV LKRPVYVREYLSPLNSRQDVHLYKGMATYFALNRTQARLPEVLSNLAQRG MKRLTSTRL"

7032..7132

repeat_region

/note="12 repeat: matches 79..206 of consensus"

7096..7098

variation

/gene="RPS6KA2"

/note="clone 427A4; TCT in this entry; substitution"

/replace="ttc"

8297..8299

variation

/gene="RPS6KA2"

/note="clone 427A4; ATA in this entry; substitution"

/replace="aca"

8349..8351

variation

/gene="RPS6KA2"

/note="clone 427A4; GTG in this entry; substitution"

/replace="gcg"

8781..8862

repeat_region

/note="41 copies 2 mer gt 72% conserved"

8788..8859

repeat_region

/note="6 copies 12 mer 81% conserved"

8796..8863

repeat_region

/note="17 copies 4 mer tgtg 78% conserved"

8870..9081

repeat_region

/note="106 copies 2 mer gt 60% conserved"

8875..9082

repeat_region

/note="52 copies 4 mer tgtg 60% conserved"

8877..9548

repeat_region

/note="16 copies 42 mer 55% conserved"

```

repeat_region      8937..9128
                    /note="6 copies 32 mer 66% conserved"
repeat_region      8979..9538
                    /note="10 copies 56 mer 57% conserved"
repeat_region      9203..9554
                    /note="11 copies 32 mer 58% conserved"
repeat_region      9205..9554
                    /note="5 copies 70 mer 63% conserved"
repeat_region      9234..9482
                    /note="3 copies 83 mer 71% conserved"
repeat_region      9334..9533
                    /note="50 copies 4 mer 91gt 58% conserved"
repeat_region      9337..9544
                    /note="4 copies 52 mer 69% conserved"
misc_feature        /note="4 copies 52 mer 69% conserved"
                    /complement(10081..>10287)
repeat_region      /note="match: EST: Em:AA601621"
                    10710..10867
misc_feature        /note="2 copies 79 mer 96% conserved"
                    /complement(10959..11359)
misc_feature        /note="match: GSS: Em:AG614575"
                    /complement(10962..11370)
misc_feature        /note="match: GSS: Em:AG614609"
                    11361..11794
misc_feature        /gene="RPS6KA2"
                    /note="match: GSS: Em:AQ268269"
                    11933..12432
misc_feature        /gene="RPS6KA2"
                    /note="match: GSS: Em:AQ128333"
                    12623..13623
misc_feature        /gene="RPS6KA2"
                    /note="match: STS: Em:HS427A4T"
                    13688..14008
repeat_region      /note="MER46B repeat: matches 1..233 of consensus"
                    14450..14664
misc_feature        /gene="RPS6KA2"
                    /note="match: GSS: Em:AQ036341"
                    14665..15015
repeat_region      /note="LIMB3 repeat: matches 5816..6184 of consensus"
                    16306..16449
repeat_region      /note="2 copies 72 mer 90% conserved"
                    16591..16980
repeat_region      /note="MTLD repeat: matches 84..505 of consensus"
                    17132..17492
repeat_region      /note="THE1B repeat: matches 1..362 of consensus"
                    20519..20832
repeat_region      /note="AluX repeat: matches 3..312 of consensus"
                    22162..23991
misc_feature        /gene="RPS6KA2"
                    /note="match: ESTs: Em:F11266 Em:F08927"
                    23013..23202
repeat_region      /note="MER20 repeat: matches 1..218 of consensus"
                    23872..24197
repeat_region      /note="AluY repeat: matches 1..311 of consensus"

Query Match      1.5%; Score 21; DB 9; Length 178098;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 930 attgttaataatttcac 950
      |||
Db 136272 ATTGTTAAATATTTTCAA 136292

```

```

RESULT 15
AC009480      188353 bp      DNA      PRI      09-MAY-2001
LOCUS      AC009480      Homo sapiens BAC clone RP11-295N18 from 2, complete sequence.
DEFINITION      AC009480      Homo sapiens BAC clone RP11-295N18 from 2, complete sequence.
ACCESSION      AC009480.4      GI:11181845
VERSION      HTG.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS      Sulston,J.E. and Waterston,R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
REFERENCE
AUTHORS      2 (bases 1 to 188353)
TITLE      Lamar,B., Drone,K., Laplant,Y. and Gibson,A.
JOURNAL      The sequence of Homo sapiens BAC clone RP11-295N18
JOURNAL      Unpublished
JOURNAL      3 (bases 1 to 188353)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (24-AUG-1999) Genome Sequencing Center, Washington
JOURNAL      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
REFERENCE
AUTHORS      4 (bases 1 to 188353)
TITLE      Waterston,R.H.
JOURNAL      Direct Submission
JOURNAL      Submitted (16-NOV-2000) Genome Sequencing Center, Washington
JOURNAL      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
REFERENCE
AUTHORS      5 (bases 1 to 188353)
TITLE      Waterston,R.
JOURNAL      Direct Submission
JOURNAL      Submitted (09-MAY-2001) Department of Genetics, Washington
JOURNAL      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL      On Nov 16, 2000 this sequence version replaced gi:7630931.
COMMENT
JOURNAL      Genome Center
JOURNAL      Center: Washington University Genome Sequencing Center
JOURNAL      Center code: WUGSC
JOURNAL      Web site: http://genome.wustl.edu/gsc
JOURNAL      Contact: saplens@watson.wustl.edu
JOURNAL      Summary Statistics
JOURNAL      Center project name: H_NEO295N18

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Firenzen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is RP11-567F11, 200 bp overlap; the clone sequenced to the right is CTD-2011N5. Actual start of this clone is at base position 195 of RP11-295N18; actual end is at base position 188353 of RP11-295N18.

FEATURES

Source

1..188353

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-295N18"
/clone_id="RPC1-11"
475..812
/rpt_family="MER2_type"
repeat_region
/rpt_family="Alu"
1776..2046
/rpt_family="Alu"
2057..2349
/rpt_family="Alu"
3373..3766
/rpt_family="MALR"
4300..4609
/rpt_family="Alu"
5226..5508
/rpt_family="L1"
5507..6008
/rpt_family="L1"
10521..10838
/rpt_family="L2"
11126..11176
/rpt_family="L2"
11695..11825
/rpt_family="Alu"
11864..12591
/rpt_family="L1"
12894..13111
/rpt_family="MIR"
13773..13957
/rpt_family="L1"
14043..14110
/rpt_family="Mariner"
14263..14475
/rpt_family="L1"
14466..14612
/rpt_family="L1"
14508..14759
/rpt_family="L1"
14760..14900
/rpt_family="L1"
15073..15197
/rpt_family="MER53"
16391..16556
/rpt_family="MIR"
17403..17742
/rpt_family="MER1_type"
18054..18355
/rpt_family="Alu"
20139..20224
/rpt_family="MER1_type"
20258..20581
/rpt_family="MER2_type"
22584..23537
/rpt_family="L1"
23538..24237
/rpt_family="L1"
28743..29952
/rpt_family="MIR"
30092..30204
/rpt_family="CR1"
30213..30368
/rpt_family="MER1_type"
31615..31802
/rpt_family="MIR"
33760..34070
/rpt_family="Alu"
34251..34295
/rpt_family="CR1"
34380..34499
/rpt_family="CR1"
37237..37286
repeat_region

/rpt_family="MIR"
36439..38599
/rpt_family="MIR"
40806..40879
/rpt_family="CR1"
45154..45465
/rpt_family="Alu"
45703..45795
/rpt_family="L2"
45953..46125
/rpt_family="L2"
47026..47350
/rpt_family="Alu"
47413..47523
/rpt_family="MIR"
47524..48271
/rpt_family="L1"
48272..48387
/rpt_family="MIR"
49668..49707
/rpt_family="L1"
51195..51294
/rpt_family="L2"
51310..51432
/rpt_family="MER105"
52841..52974
/rpt_family="MER1_type"
52864..53002
/rpt_family="MER1_type"
53069..53120
/rpt_family="MIR"
57789..58078
/rpt_family="Alu"
59027..59157
/rpt_family="MER1_type"
59161..59409
/rpt_family="MER2_type"
59475..60200
/rpt_family="MER2_type"
60251..60774
/rpt_family="MER1_type"
60775..62208
/rpt_family="L1"
62214..62658
/rpt_family="MER1_type"
62825..63134
repeat_region

Query Match 1.5%; Score 21; DB 9; Length 188353;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1194 aaggtggttcaagaatgg 1214
Db 98676 AAGGTGCTTACCAAGATGG 98696

RESULT 16
AL592214 195277 bp DNA HTG 02-AUG-2001
LOCUS Mus musculus chromosome 1 clone RP23-212174, *** SEQUENCING IN
DEFINITION PROGRESS ***, In unordered pieces.
ACCESSION AL592214
VERSION ~ AL592214.3 GI:14588776
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 195277)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone

```

requests: clonerequest@sanger.ac.uk
On Jul 3, 2001 this sequence version replaced gi:14586345.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM212124
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 193820 bases at least Q40
Consensus quality: 194186 bases at least Q30
Consensus quality: 194454 bases at least Q20
Insert size: 194977; sum-of-contigs
Insert size: 186154; 8.1% error; agarose-fp
Quality coverage: 8.50x in Q20 bases; sum-of-contigs Quality
coverage: 8.90x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

Location/Qualifiers
source 1. 195277
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="RP23-212124"
/clone_11b="RPCT-23"
1. 43565
/note="assembly_fragment:04211"
43666. 93728
/note="assembly_fragment:03891"
fragment_chain:1
93829. 154402
/note="assembly_fragment:00954"
fragment_chain:1
154503. 195277
/note="assembly_fragment:02416"
fragment_chain:1
clone_end:SP6
vector_side:right

BASE COUNT 63064 a 35235 c 35934 g 60739 t 305 others
ORIGIN
Query Match 1.5%; Score 21; DB 2; Length 195277;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 231 tgaacagagagccttggaat 251
|||||
Db 159862 TGAACAGAGCCTTGGAAAT 159882

RESULT 17
AC021626
LOCUS AC021626 196788 bp DNA HTG 12-MAR-2000
DEFINITION Homo sapiens clone Rpl1-39112, WORKING DRAFT SEQUENCE, 28 unordered
pieces.
AC021626
VERSION AC021626.3 GI:7229839
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 196788)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone Rpl1-39112

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 196788)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P.,
Darellano,K., Dewar,K., Domino,M., Doyle,M., Fenster,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Klen,C., Karatas,A., Klein,J.,
Landers,J., Leoccky,J., Levine,R., Liu,C., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Poyre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6978227.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: 15883
Center clone name: 391-1-2

Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 181094 bases at least Q40
Consensus quality: 189581 bases at least Q30
Consensus quality: 192532 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 194088; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1281: contig of 1281 bp in length
* 1282 1381: gap of 100 bp
* 1382 2899: contig of 1518 bp in length
* 2900 2999: gap of 100 bp
* 3000 5681: contig of 2682 bp in length
* 5682 5781: gap of 100 bp
* 5782 8351: contig of 2570 bp in length
* 8352 8451: gap of 100 bp
* 8452 10639: contig of 2188 bp in length
* 10640 10739: gap of 100 bp
* 10740 13653: contig of 2916 bp in length
* 13656 13753: gap of 100 bp
* 13756 15927: contig of 2172 bp in length
* 15928 16027: gap of 100 bp
* 16028 19028: contig of 3001 bp in length
* 19029 19128: gap of 100 bp
* 19129 22669: contig of 3541 bp in length
* 22670 22769: gap of 100 bp
* 22770 26791: contig of 4022 bp in length

```

* 26792 26891: gap of 100 bp
* 26892 30845: contig of 3954 bp in length
* 30846 30945: gap of 100 bp
* 30946 35594: contig of 4649 bp in length
* 35595 35694: gap of 100 bp
* 35695 40148: contig of 4454 bp in length
* 40149 40248: gap of 100 bp
* 40249 45042: contig of 4794 bp in length
* 45043 45142: gap of 100 bp
* 45143 49477: contig of 4335 bp in length
* 49478 49577: gap of 100 bp
* 49578 54518: contig of 4941 bp in length
* 54519 54618: gap of 100 bp
* 54619 61982: contig of 7364 bp in length
* 61983 62082: gap of 100 bp
* 62083 67811: contig of 5729 bp in length
* 67812 67911: gap of 100 bp
* 67912 74070: contig of 6159 bp in length
* 74071 74170: gap of 100 bp
* 74171 80685: contig of 6515 bp in length
* 80686 80785: gap of 100 bp
* 80786 88056: contig of 8271 bp in length
* 88057 88156: gap of 100 bp
* 88157 99104: contig of 9948 bp in length
* 99105 99204: gap of 100 bp
* 99205 109014: contig of 9810 bp in length
* 109015 109114: gap of 100 bp
* 109115 122471: contig of 13357 bp in length
* 122472 122571: gap of 100 bp
* 122572 138753: contig of 16182 bp in length
* 138754 138853: gap of 100 bp
* 138854 156691: contig of 17838 bp in length
* 156692 156791: gap of 100 bp
* 156792 173620: contig of 16829 bp in length
* 173621 173720: gap of 100 bp
* 173721 196788: contig of 23068 bp in length.

```

```

FEATURES
    source
        1..196788
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_id="RPCT-11 Human Male BAC"

```

```

misc_feature
    /note="assembly-fragment"
    1..1281
misc_feature
    /note="assembly-fragment"
    1382..2859
misc_feature
    /note="assembly-fragment"
    3000..5681
misc_feature
    /note="assembly-fragment"
    5782..8351
misc_feature
    /note="assembly-fragment"
    8452..10639
misc_feature
    /note="assembly-fragment"
    10740..13655
misc_feature
    /note="assembly-fragment"
    13756..15927
misc_feature
    /note="assembly-fragment"
    clone_end:77
vector_side:right
    16028..19028
misc_feature
    /note="assembly-fragment"
    19129..22669
misc_feature
    /note="assembly-fragment"
    22770..26791
misc_feature
    /note="assembly-fragment"
    26892..30845
misc_feature
    /note="assembly-fragment"
    30946..35594
misc_feature
    /note="assembly-fragment"
    35695..40148
misc_feature
    /note="assembly-fragment"
    40249..45042
misc_feature
    /note="assembly-fragment"
    45143..49477

```

```

misc_feature
    /note="assembly-fragment"
    49578..54518
    /note="assembly-fragment"
    54619..61982
    /note="assembly-fragment"
    62083..67811
    /note="assembly-fragment"
    67912..74070
    /note="assembly-fragment"
    74171..80685
    /note="assembly-fragment"
    80786..88056
    /note="assembly-fragment"
    88157..99104
    /note="assembly-fragment"
    99205..109014
    /note="assembly-fragment"
    109115..122471
    /note="assembly-fragment"
    122572..138753
    /note="assembly-fragment"
    138854..156691
    /note="assembly-fragment"
    156792..173620
    /note="assembly-fragment"
    clone_end:526
vector_side:right
    173721..196788
misc_feature
    /note="assembly-fragment"
    53672 a 43305 c 43403 g 53703 t 2705 others
BASE COUNT
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 21; DB 2; Length 196788;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 18
LOCUS AE003487/c
DEFINITION Drosophila melanogaster genomic scaffold 142000013386053 section 4
ACCESSION AE003487 AE002593
VERSION AE003487.1 GI:7292680
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 301051)
Adams,M.D., Celnikher,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandal,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,R.H., Blazer,R.G., Chame,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Miklos,G.L., Abiri,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Bailew,R.M., Beeson,K.Y.,
Baxendale,J., Bayraktaroglu,U., Beasley,E.M., Beeson,K.Y.,
Bencos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Borchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Deichner,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,

```

gene	protein_id="AAFA48079.1" /db_xref="GI:7292682" /translation="MLYRQLFGASANSNARALLIFVLAINGLYVYVAVLTWQGO AGSELNRRISQALQIHERROENESLKQIDELASDQIDKRSAKVLQRENDL NKRLEPAVGEPESEMFSPADLDRGNVNAEGAPNDIEAVDPHGSEFSELEFTR RATQIONGINNFESELEGKVRAAVAAASDADEESINQVLQGAHRSKLSLDMR MOSQGYEMRHRKEARDLSDVORLHLQAPSDQNNRKYCLANKRSKSLDMR VYCFIVAYTERTLILKRSNGYHKGGEVEYFQPSNCSHAGTANYNPGKPTQV LVPIIDSLMPPEPLPLPAVPEDLAPLRKLRHGPIVWVGQFLYLRPOPTRODL TSGMRINQWERPIVAVHRNDKQTEACGASVEEMTVYDRTYRTLENGSFTARI FLASDAQVIEPARRKYXOYOLGPEAPARASASVEEMTVYDRTYRTLENGSFTARI LVCTTSSQYCRVAYIKOMTPDPAHRKSLDDITYGQANRRRYIAKRPETHD LDLRGDLVSVAQNWDSKSKNTKRGQGLFSPFKVEEVYDAKPLRAGT complement(join(7627. .8483,8886. .9925,10592. .>10631)) /gene="CG2446" /product="Ct43088" /db_xref="FLYBASE:FBgn0030328" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10139. .>10310)) /gene="CG2446" /product="Ct43627" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG2446" /product="Ct8088" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .>11123)) /gene="CG2446" /map="10D4-10D5" /db_xref="FLYBASE:FBan0002446" /db_xref="FLYBASE:FBgn0030328" complement(join(7627. .8483,8886. .9805,9872. .9925, 10675. .>11123)) /gene="CG244
------	--

/note="CG2446 gene product [alt 3]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0002446"
 /db_xref="FLYBASE:FBgn0030328"
 /protein_id="AAF48080.1"
 /db_xref="GI:7292683"
 /translation="MSNGKATVFFFGSTKPEYCYOLYPOVLKLAERCKKPPQEL
 IRLDQWYNEELKIKARGKDAHVYELVQSMKQSRGKFPOLSLYKVTNPRAV
 IOETKARKEIPNLEQATFALSNKGVCTTASALLAAAPDSAPMAADCIAPPI
 EGIDITKEITLVNHIQATVERLNAEYGGTTPASHRELALSHIVADLSPEKL
 DDMPPGSGASTGSLSTNGNSKVLDDGTNDGVLDDEDSQAGRTATSEET
 ENENTNPALPLQSGEAKNNAAVGAALQGDSENFVNSDTSQEPIDNDGTTQT
 ATTSDEDEPIALDIGIGSSGTPLASDSNDEAPKTSILPITPHSSNQMO
 KQSPQPKTNSTNNQOPAPLAEEEVTAPOPAKATAPANGNGNGVYGDSD
 EDPADEDEDELDEEDNEALEADESSNSGIYRDSKLOOLANKAVDAVSPVAAAG
 DSPATIGOKRTALHCDMELKNAGGVGVCGRKSPDLKRESE"
 join(13155..13381,14485..14887,14949..16011)
 /gene="CG1837"
 /product="CT5608"
 /db_xref="FLYBASE:FBan0001837"
 /db_xref="FLYBASE:FBgn0030329"
 <13155..>16011
 /gene="CG1837"
 /map="10D4-10D6"
 /db_xref="FLYBASE:FBan0001837"

Query Match 1.5%; Score 21; DB 3; Length 301051;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 520 gttgcattgaacaattagct 540
 |||||||||||||||||||
 Db 221393 gttgcattgaacaattagct 221373

RESULT 19
 LOCUS AX087938 32 bp DNA
 DEFINITION Sequence 4 from Patent WO0114535.
 ACCESSION AX087938 PAT 17-MAR-2001
 VERSION AX087938.1 GI:13396916
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 32)
 AUTHORS Schwientek,T. and Clausen,H.
 TITLE Udp-n-acetylglucosamine: galactose-g(b)1,3-n
 -acetylglucosamine-g(a)-r/(glcnac to galnac)-g(b)1,6-n
 -acetylglucosaminyltransferase e, c2gnl3
 Patent: WO 0114535-A 4 01-MAR-2001;
 JOURNAL Schwientek, Tilo (DK); Clausen, Henrik (DK)
 FEATURES
 source location/Qualifiers
 1..32
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="a primer named TSHC99"
 BASE COUNT 13 a 4 c 7 g 8 t
 ORIGIN

Query Match 1.5%; Score 20; DB 6; Length 32;
 Best Local Similarity 100.0%; Pred. No. 8.9;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgaagatatcaatgtta 20
 |||||||||||||||||||
 Db 13 atgaagatatcaatgtta 32
 RESULT 20
 HUMPSABD/c
 LOCUS HUMPSABD 173 bp DNA PRI 27-APR-1993

DEFINITION Human protein S beta pseudogene, exon n+3.
 ACCESSION M57314.102919
 VERSION M57314.1 GI:190562
 KEYWORDS protein S; pseudogene; sex hormone binding globulin-like.
 SOURCE Human DNA.
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 173)
 AUTHORS Edestrand,C.W., Lundwall,A., Wydro,R. and Stenflo,J.
 TITLE Molecular analysis of the gene for vitamin K dependent protein S
 and its pseudogene. Cloning and partial gene organization
 JOURNAL Biochemistry 29, 7861-7868 (1990)
 MEDLINE
 FEATURES
 source location/Qualifiers
 1..173
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 55 a 26 c 42 g 50 t
 ORIGIN
 Map position 3p11-3q11.2.

Query Match 1.5%; Score 20; DB 9; Length 173;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 901 ggcagtcattattgttt 920
 |||||||||||||||||||
 Db 105 GGCAGTGTATTGTTT 86

RESULT 21
 LOCUS HUMPSB11/c 185 bp DNA PRI 08-JAN-1995
 DEFINITION Human protein S pseudogene beta (PS-beta), exon 11.
 ACCESSION M36575.102918
 VERSION M36575.1 GI:190319
 KEYWORDS S protein.
 SEGMENT 11 of 12
 SOURCE Human DNA.
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 185)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Ploos van Amstel,H.K., Reltsma,P.H., van der Logt,C.P. and
 Bertina,R.M.
 JOURNAL Intron-exon organization of the active human protein S gene PS
 alpha and its pseudogene PS beta: duplication and silencing during
 primate evolution
 Biochemistry 29 (34), 7853-7861 (1990)
 MEDLINE
 COMMENT Draft entry and computer-readable sequence for Biochemistry 29,
 7853-7861 (1990) kindly submitted
 by H.K.Ploos van Amstel, 13-JUL-1990.
 FEATURES
 source location/Qualifiers
 1..185
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="3p21-cen"
 prim_transcript <1..>185
 /gene="PROS2"
 /note="protein S pseudogene beta mRNA and introns"
 <1..10
 /gene="PROS2"
 /note="intron 11"
 <11..179
 /gene="PROS2"
 /note="protein S pseudogene beta, 1; G00-120-757"
 /pseudo
 /number=11
 11..179
 /gene="PROS2"
 exon

/note="Protein S pseudogene beta, exon 11; G00-120-757;
putative"
180. >185
/gene="PROS2"
/note="Intron 12"
BASE COUNT 59 a 26 c 44 g 56 t
ORIGIN

Query Match 1.5%; Score 20; DB 9; Length 185;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 901 ggcagtcctatttgcgttt 920
DB 113 ggcagtcctatttgcgttt 94

RESULT 22
HMPSPS08/c 478 bp DNA PRI 10-JAN-1995
LOCUS HMPSPS08
DEFINITION Human protein S-beta pseudogene, exon 12.
ACCESSION M57861.1 J02917
VERSION M57861.1 GI:190682
KEYWORDS protein S; pseudogene.
SEGMENT 8 of 11
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 478)
AUTHORS Schmidt, D.K., Taito, A.V., Phelps, L.G., Tomczak, J.A. and Long, G.L.
TITLE Organization of the human protein S genes
JOURNAL Biochemistry 29 (34), 7845-7852 (1990)
MEDLINE 9108444

FEATURES
source
1. 478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/feature_type="CDS"
/map="3p21-cent"
order(M57861.1:345..532.1..95)
/gene="PS-beta"
/number=11
exon
96..264
/note="PROS2"
/number=12
BASE COUNT 151 a 71 c 78 g 178 t
ORIGIN

Query Match 1.5%; Score 20; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 901 ggcagtcctatttgcgttt 920
DB 198 ggcagtcctatttgcgttt 179

RESULT 23
MUSCYP32 732 bp DNA ROD 27-APR-1993
LOCUS MUSCYP32
DEFINITION Mouse cytochrome c pseudogene MC3.
ACCESSION K03126
VERSION K03126.1 GI:192872
KEYWORDS cytochrome; cytochrome c; pseudogene.
SEGMENT 2 of 2
SOURCE Mouse (BALB/c) 12 day old embryo DNA (Library of P. Leder), clone MC3.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 732)
AUTHORS Limbach, K.J., and Wu, R.
TITLE Characterization of a mouse somatic cytochrome c gene and three
JOURNAL Nucleic Acids Res. 13, 617-630 (1985)
MEDLINE 85215501
FEATURES
source
1. 732
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 268 a 136 c 160 g 168 t
ORIGIN About 310 bp after segment 1.

Query Match 1.5%; Score 20; DB 10; Length 732;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 262 ctggaataagaagaaggga 281
DB 352 CTGGAATAAGAGAAGGGA 371

RESULT 24
FLBHAZD/c 1086 bp ss-RNA VRL 02-AUG-1993
LOCUS FLBHAZD
DEFINITION Influenza B/India/3/89 haemagglutinin (HA) gene (segment 4, HA1 domain), 5' end.
ACCESSION M65168
VERSION M65168.1 GI:325155
KEYWORDS haemagglutinin; receptor-binding protein; transmembrane protein.
SOURCE Influenza virus type B (individual isolate B/India/3/89) RNA.
ORGANISM Influenza B virus
Virus; ssRNA negative-strand viruses; Orthomyxoviridae; Influenza virus A and B group.
REFERENCE 1 (bases 1 to 1086)
AUTHORS Rota, P.A., Hemphill, M., Whistler, T., Regnery, H.L., and Kendal, A.
TITLE Antigenic and genetic characterization of the haemagglutinins of recent co-circulating strains of Influenza B virus
JOURNAL J. Gen. Virol. 73, 2737-2742 (1992)
MEDLINE 93019032
FEATURES
source
1. 1086
/organism="Influenza B virus"
/isolate="B/India/3/89"
/db_xref="taxon:11520"
/map="segment 4"
1. 1086
/gene="HA"
1. >1086
/note="HA1/HA2 cleavage site is at position 1086"
/codon_start=1
/product="hemagglutinin"
/protein_id="A8443706.1"
/db_xref="GI:325156"
/translation="MKAIIVLVVTSNADRTGTTSSSPHYVKTATGCEVNTGYPVTTPKSHFANLKGKTRGKLCPCVLCCTDVALAPKCTGTPSKASILEV KPVTCGEPIMHDKRIKIRQLNLRGIEHRIKSTHNVINAEIAPGFGYKIGTSRSPN VTNGGFEATMAVAAPKNDNNKATNPILVEVPICTEGEDQITVWGFNDNETOMWK LYDNGKDKFTSSANGVTTHTVYQIGFPPDAEDGGLPQSGRIYVDIMVOKSGKGTI TYRGDILPKQVWCASGRSVKISLLEIGADDLHKRYGKLKSKRYTGEMAKKIG NCPHWKTPFLANGTIRPPAKIKER"

gene
CDS
1. 1086
/gene="HA"
1. >1086
/note="HA1/HA2 cleavage site is at position 1086"
/codon_start=1
/product="hemagglutinin"
/protein_id="A8443706.1"
/db_xref="GI:325156"
/translation="MKAIIVLVVTSNADRTGTTSSSPHYVKTATGCEVNTGYPVTTPKSHFANLKGKTRGKLCPCVLCCTDVALAPKCTGTPSKASILEV KPVTCGEPIMHDKRIKIRQLNLRGIEHRIKSTHNVINAEIAPGFGYKIGTSRSPN VTNGGFEATMAVAAPKNDNNKATNPILVEVPICTEGEDQITVWGFNDNETOMWK LYDNGKDKFTSSANGVTTHTVYQIGFPPDAEDGGLPQSGRIYVDIMVOKSGKGTI TYRGDILPKQVWCASGRSVKISLLEIGADDLHKRYGKLKSKRYTGEMAKKIG NCPHWKTPFLANGTIRPPAKIKER"

BASE COUNT 387 a 250 c 224 g 225 t
ORIGIN

Query Match 1.5%; Score 20; DB 14; Length 1086;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 763 gtgaacccccaacagtaa 782


```

|||||
Db      627 GTGAACCCCAACAGCTAA 608

RESULT 25
LOCUS   EAU59304      2165 bp      DNA      BCT      20-MAR-1997
DEFINITION
Enterobacter agglomerans chitinase Chla_Entag (chla) gene, complete
cgs.
ACCESSION
VERSION  U59304.1      GI:1899047
KEYWORDS
SOURCE    Pantoea agglomerans.
           Pantoea agglomerans.
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Pantoea.
REFERENCE
AUTHORS   1 (bases 1 to 2165)
           Chernin, L.S., De la Fuente, L., Sobolev, V., Haran, S., Vorgias, C.E.,
           Oppenheim, A.B. and Chet, I.
TITLE      Molecular cloning, structural analysis, and expression in
           Escherichia coli of a chitinase gene from Enterobacter agglomerans
           Appl. Environ. Microbiol. 63 (3), 834-839 (1997)
JOURNAL    97208202
MEDLINE
REFERENCE  2 (bases 1 to 2165)
           Chernin, L.S., de la Fuente, L., Sobolev, V., Ismailov, Z., Haran, S.,
           Vorgias, C.E., Oppenheim, A.B. and Chet, I.
AUTHORS
TITLE      Direct Submission
           Submitted (27-MAY-1996) Plant Pathology and Microbiology, Hebrew
           University of Jerusalem, Faculty of Agriculture, Rehovot 76100,
           Israel
FEATURES
source     1. 2165
           /organism="Pantoea agglomerans"
           /strain="IC1270"
           /db_xref="taxon:549"
           28. 1716
           /gene="chla"
           28. 1716
           /EC_number="3.2.1.14"
           /codon_start=1
           /transl_table=1
           /product="Chitinase Chla_Entag"
           /protein_id="AAB49933.1"
           /db_xref="GI:1899048"
           /translation="MRKFNKPLALLIGSLCSAQAARAGKPSIASGPKFAIVEYD
           OQATYANNIVKTKTAADYSTSMDSGDAGTAKVILDGKVEWSSGATGSGANPKY
           NKGRTQMOVALCNADCTASDTQIYATDGSRLAPLEPLENNKPKKDSCKYV
           GSTFVGVGKRNFTVDKTPAONITLILYGFIPVCGDGLNDSKVGESFALQKSC
           QGRDFKVSJHDPFAAVQKQKVTAWDDPKGFGQMLAKQARPDLKILPSIGWT
           LSDPFFMGKRVKRDPRVGSYKEFLQTKFFEDVDIMEPFGGAGNPKILGNQDGT
           YVQMLKDLRALMDQLSAETGKRYELTSAISGKRIKRVYNTAQNMDIIFLMSYDF
           YGAFDLKLGHOITLAKPGRHGLTGERGROCAAGGVGPKIVYAAKYGKMGV
           SGYONNNPFTGTATGPKYKWTGNIYVROIARFISIDEMOVSADAAEAPYKPEST
           GDLTPDPDRVQAKKXIVLDKQGGCFSEHIDANDNIDILNNTSLGNSAGAQ"
BASE COUNT
ORIGIN     543 a      534 c      595 g      493 t

Query Match      1.5%; Score 20; DB 1; Length 2165;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 ttcatcctgttttaacc 67
|||||
Db 1861 TTTCATCTCTTTTAAACC 1880

RESULT 26
LOCUS   AF327184S1      2705 bp      DNA      ROD      11-JAN-2001
DEFINITION
Mus musculus inducible costimulatory protein (Icos) gene, exon 1.
ACCESSION
AF327184

```

```

VERSION   AF327184.1      GI:12083055
KEYWORDS
SEGMENT   1 of 2
SOURCE    house mouse.
ORGANISM  Mus musculus.
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS   1 (bases 1 to 2705)
           McAdam, A.J., Greenwald, R.J., Levin, M.A., Chernova, T.,
           Malenkovich, N., Ling, V., Freeman, G.J. and Sharpe, A.H.
TITLE      The inducible costimulatory (ICOS) molecule is critical for
           CD40-mediated antibody class switching
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 2705)
           McAdam, A.J., Greenwald, R.J., Levin, M.A., Chernova, T.,
           Malenkovich, N., Ling, V., Freeman, G.J. and Sharpe, A.H.
AUTHORS
TITLE      Direct Submission
           Submitted (08-DEC-2000) Adult Oncology, Dana-Farber Cancer
           Institute, 44 Binney St. Boston, MA 02115, USA
FEATURES
source     Location/Qualifiers
           1. 2705
           /organism="Mus musculus"
           /strain="129/Ola"
           /db_xref="taxon:10090"
           <1382..1439
           /gene="Icos"
           /number=1
           exon

BASE COUNT      836 a      537 c      446 g      886 t
ORIGIN

Query Match      1.5%; Score 20; DB 10; Length 2705;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 gtgacattatcacagctcta 348
|||||
Db 2146 GTGACATTATTCAGACTCTA 2165

RESULT 27
LOCUS   AL391815/c      61274 bp      DNA      PRI      09-APR-2001
DEFINITION
Human DNA sequence from clone RP11-223L24 on chromosome 6 Contains
STSs and GSSs, complete sequence.
ACCESSION   AL391815
VERSION     AL391815.6      GI:11121081
KEYWORDS
SOURCE      Human.
ORGANISM    Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 61274)
           Williams, S.
AUTHORS
TITLE      Direct Submission
           Submitted (28-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
           requests: clonerequests@sanger.ac.uk
           On Nov 8, 2000 this sequence version replaced gi:10933315.
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submission
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
           Group. Further information can be found at

```

http://www.sanger.ac.uk/HGP/Chr6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-223L24. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-223L24 is at 1 in this sequence.
 The true right end of clone RP11-223L24 is at 61175 in this sequence.
 The true left end of clone RP1-90L14 is at 46903 in this sequence.
 The true right end of clone RP1-153N22 is at 46903 in this sequence.
 This sequence was finished as follows unless otherwise
 noted: all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.
 RP11-223L24 is from the library RP11-1.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/dacpac/home.htm
 VECTOR: pBACe3.6

FEATURES

source

1. 61274
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-223L24"
 /clone_1fb="RP11-11.1"
 complement(966..1493)
 misc_feature
 /note="match: GSS: Em:AQ677601"
 1429..1571
 repeat_region
 /note="MIR repeat: matches 113..261 of consensus"
 1620..1758
 repeat_region
 /note="MIR repeat: matches 47..198 of consensus"
 complement(2962..3318)
 misc_feature
 /note="match: GSS: Em:AQ33701"
 complement(3212..3661)
 /note="match: GSS: Em:AQ673685"
 3242..3309
 repeat_region
 /note="2 copies 34 mer 91% conserved"
 3345..3849
 misc_feature
 /note="match: GSS: Em:AQ336206"
 4101..4136
 repeat_region
 /note="L1P45 repeat: matches 6108..6143 of consensus"
 4591..4835
 repeat_region
 /note="L1P83 repeat: matches 5895..6138 of consensus"
 5358..5571
 repeat_region
 /note="MIR repeat: matches 7..262 of consensus"
 complement(5536..5850)
 misc_feature
 /note="match: GSS: Em:AQ776450"
 5945..6051
 repeat_region
 /note="MIR5B repeat: matches 1..110 of consensus"
 6283..6504
 repeat_region
 /note="AluJo repeat: matches 76..288 of consensus"
 7050..7471
 repeat_region
 /note="MSTA repeat: matches 1..426 of consensus"
 9416..9861
 misc_feature
 /note="match: GSS: Em:AQ03676"
 10177..10227
 repeat_region
 /note="L2 repeat: matches 2688..2738 of consensus"
 11006..11158
 repeat_region
 /note="L1M1 repeat: matches 6105..6262 of consensus"
 12018..12195
 repeat_region
 /note="MIR repeat: matches 49..252 of consensus"
 12491..12545
 repeat_region
 /note="MIR1 repeat: matches 309..364 of consensus"
 12708..12842
 repeat_region
 /note="MIR11 repeat: matches 17..157 of consensus"
 12811..12857
 repeat_region
 /note="MIR1J repeat: matches 3..51 of consensus"
 13579..13718
 repeat_region
 /note="L2 repeat: matches 2211..2356 of consensus"
 13897..14157
 misc_feature
 /note="match: GSS: Em:AQ564337"
 14320..14496
 repeat_region
 /note="MIR5B repeat: matches 2..177 of consensus"

repeat_region
 14569..14645
 /note="L2 repeat: matches 2620..2713 of consensus"
 15346..15646
 repeat_region
 /note="AluX repeat: matches 1..303 of consensus"
 18563..18722
 repeat_region
 /note="AluJo/FRM repeat: matches 146..307 of consensus"
 18794..19307
 repeat_region
 /note="MIR1D repeat: matches 1..505 of consensus"
 complement(19511..19649)
 misc_feature
 /note="match: STS: Em:G07838"
 19523..19645
 misc_feature
 /note="match: GSS: Em:A2046652 Em:A2126672"
 complement(19527..19655)
 misc_feature
 /note="match: GSS: Em:A249601"
 complement(19534..19636)
 misc_feature
 /note="match: STS: Em:G08798"
 complement(19534..19632)
 misc_feature
 /note="match: GSS: Em:A2103542"
 complement(19541..19636)
 misc_feature
 /note="match: GSS: Em:AQ683084"
 complement(19542..19656)
 misc_feature
 /note="match: GSS: Em:A2020257"
 complement(19543..19671)
 misc_feature
 /note="match: STS: Em:G12422 Em:G15742"
 complement(19543..19656)
 misc_feature
 /note="match: GSS: Em:A2040015 Em:A2040457"
 19544..19651
 misc_feature
 /note="match: STS: Em:G15739"
 complement(19544..19656)
 misc_feature
 /note="match: GSS: Em:A2122363 Em:A2271663
 match: STS: Em:L16408"
 19544..19699
 misc_feature
 /note="match: GSS: Em:A2446995"
 19544..19663
 misc_feature
 /note="match: GSS: Em:AQ52676 Em:AQ528204"
 19544..19656
 misc_feature
 /note="match: GSS: Em:A2318431 Em:A2370392"
 complement(19544..19639)
 misc_feature
 /note="match: GSS: Em:B51766"
 19544..19640
 misc_feature
 /note="match: STS: Em:G27274"
 complement(19546..19645)
 misc_feature
 /note="match: STS: Em:L17665"
 complement(19546..19658)
 misc_feature
 /note="match: STS: Em:G08426"
 19547..19644
 misc_feature
 /note="match: GSS: Em:AQ379803"
 19549..19635
 misc_feature
 /note="match: STS: Em:G10237"
 complement(19549..19655)
 misc_feature
 /note="match: GSS: Em:A225034"
 complement(19552..19648)
 misc_feature
 /note="match: STS: Em:G08887"
 19816..19964
 repeat_region
 /note="MIR1 repeat: matches 1..141 of consensus"
 21610..21675
 repeat_region
 /note="33 copies 2 mer ga 69% conserved"
 21621..21880
 repeat_region
 /note="15 copies 4 mer aaag 83% conserved"
 21738..21859
 repeat_region
 /note="FIAM_A repeat: matches 8..129 of consensus"
 24497..24568
 repeat_region
 /note="18 copies 4 mer ttcc 87% conserved"
 24512..24609
 repeat_region
 /note="49 copies 2 mer ct 65% conserved"
 24656..24953
 repeat_region
 /note="AluJg repeat: matches 1..297 of consensus"
 25114..25425
 repeat_region
 /note="AluX repeat: matches 1..312 of consensus"
 25493..25726
 repeat_region
 /note="MIR repeat: matches 31..261 of consensus"
 25951..26102
 repeat_region
 /note="L1M3 repeat: matches 6004..6162 of consensus"

```

repeat_region      26257..27168
                    /note="LMC5 repeat: matches 6640..7515 of consensus"
repeat_region      27243..27608
                    /note="THEIB repeat: matches 1..364 of consensus"
repeat_region      27609..28708
                    /note="THEIB-INTERNAL repeat: matches 490..1380 of
                    consensus"
repeat_region      28709..29020
                    /note="AluX repeat: matches 1..311 of consensus"
repeat_region      29021..29508
                    /note="THEIB-INTERNAL repeat: matches 1..490 of consensus"
repeat_region      29509..29870
                    /note="THEIB repeat: matches 5..364 of consensus"
repeat_region      29871..30258
                    /note="AluSc repeat: matches 3..274 of consensus"
repeat_region      30678..31184
                    /note="LMXB4 repeat: matches 5795..6287 of consensus"
repeat_region      31378..31489
                    /note="56 copies 2 mer tt 63% conserved"
repeat_region      31492..31676
                    /note="LMXB4 repeat: matches 6007..6180 of consensus"
repeat_region      31773..33606
                    /note="LMXB4 repeat: matches 4425..6303 of consensus"

```

```

Query Match      1.5%: Score 20; DB 9; Length 61274;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      566 ctccaattagagctgtg 585
        |||||
Db      750 CTTCCAATTAGAGGCTGTG 731

```

```

RESULT 28
AL513123      70339 bp      DNA      PRI      27-Apr-2001
LOCUS      Human DNA sequence from clone RP11-346D19 on chromosome 6, complete
DEFINITION
ACCESSION      AL513123
VERSION      AL513123.10 GI:13897160
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 70339)
AUTHORS      Blakey/S.
JOURNAL      Direct Submission
              Submitted (27-Apr-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Apr 30, 2001 this sequence version replaced gi:13443468.

```

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone configs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-346D19 is from the library RPI1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm>
 VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-346D19. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-367G18 is at 70240 in this sequence. The true right end of clone RPI-124C6 is at 100 in this sequence.

FEATURES

```

source      location/Qualifiers
            1..70339
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="RP11-346D19"
            /clone_1b="RPI1-11.2"
            628..921
            /note="AluYb repeat: matches 1..306 of consensus"
            1834..2459
            /note="L2 repeat: matches 1483..2247 of consensus"
            2578..2875
            /note="AluJo repeat: matches 5..298 of consensus"
            3839..3934
            /note="MLT1-INTERNAL repeat: matches 918..1012 of
            consensus"
            4002..4201
            /note="MER5B repeat: matches 2..175 of consensus"
            4929..5186
            /note="AluSg1 repeat: matches 1..287 of consensus"
            5346..5578
            /note="L2 repeat: matches 1660..1887 of consensus"
            5781..5875
            /note="L2 repeat: matches 2171..2268 of consensus"
            5876..6188
            /note="AluY repeat: matches 1..311 of consensus"
            6189..6256
            /note="L2 repeat: matches 2268..2337 of consensus"
            6257..6550
            /note="AluSg repeat: matches 3..300 of consensus"
            6555..6839
            /note="AluSx repeat: matches 1..305 of consensus"
            6860..6971
            /note="L2 repeat: matches 2337..2438 of consensus"
            6988..7152
            /note="L2 repeat: matches 5986..6147 of consensus"
            7567..7748
            /note="L2 repeat: matches 2541..2725 of consensus"
            8539..8842
            /note="AluSc repeat: matches 1..307 of consensus"
            9810..10013
            /note="MLT1 repeat: matches 195..406 of consensus"
            10020..10099
            /note="MIR repeat: matches 1..80 of consensus"
            11344..11550
            /note="LMC/D repeat: matches 5511..5751 of consensus"
            12685..13049
            /note="MER47A repeat: matches 1..366 of consensus"
            13489..13683
            /note="MER47A repeat: matches 67..265 of consensus"
            15394..15701
            /note="AluSg repeat: matches 1..308 of consensus"
            16252..16912
            /note="L2 repeat: matches 1374..2084 of consensus"
            17367..17434
            /note="L2 repeat: matches 2682..2750 of consensus"
            19746..19884
            /note="MLT1 repeat: matches 421..541 of consensus"
            19885..19994
            /note="MER81 repeat: matches 1..114 of consensus"
            19995..20117
            /note="MLT1 repeat: matches 245..421 of consensus"

```

```

repeat_region 20677..20718
/note="21 copies 2 mer aa 76% conserved"
repeat_region 20738..21064
/note="MULTI1 repeat: matches 1..350 of consensus"
repeat_region 21144..21330
/note="MIR repeat: matches 24..240 of consensus"
repeat_region 21404..21706
/note="MIR repeat: matches 3..307 of consensus"
repeat_region 22476..22608
/note="L2 repeat: matches 2565..2707 of consensus"
repeat_region 25443..25693
/note="L1MCI repeat: matches 6067..6321 of consensus"
repeat_region 25897..27083
/note="Tigerg3b repeat: matches 1..1176 of consensus"
repeat_region 27207..27369
/note="MERSA repeat: matches 23..189 of consensus"
repeat_region 27221..27390
/note="MERSA repeat: matches 1..166 of consensus"
repeat_region 27492..27618
/note="MIR repeat: matches 122..261 of consensus"
repeat_region 28470..28713
/note="MIR repeat: matches 2..262 of consensus"
repeat_region 29425..29600
/note="MERS4 repeat: matches 1..176 of consensus"
repeat_region 31817..32196
/note="MSTA repeat: matches 1..425 of consensus"
repeat_region 32198..32245
/note="24 copies 2 mer ac 97% conserved"
repeat_region 32393..32535
/note="FLAMC repeat: matches 1..143 of consensus"
repeat_region 33417..33714
/note="AIUto repeat: matches 1..296 of consensus"
repeat_region 34739..35171
/note="AIU5g/X repeat: matches 217..307 of consensus"
repeat_region 35430..35520
/note="AIU5g/X repeat: matches 217..307 of consensus"
repeat_region 35544..35845
/note="AIUy repeat: matches 1..301 of consensus"
repeat_region 36708..36782
/note="MADE1 repeat: matches 1..76 of consensus"
repeat_region 36789..36881
/note="MERSB repeat: matches 67..177 of consensus"
repeat_region 36902..37629
/note="LIMC4 repeat: matches 6969..7648 of consensus"
repeat_region 37634..37685
/note="MERSB repeat: matches 126..178 of consensus"
repeat_region 38202..38300
/note="MERSA repeat: matches 9..106 of consensus"
repeat_region 38469..38656
/note="MERSA repeat: matches 1..184 of consensus"
repeat_region 38669..38782
/note="MERSB repeat: matches 92..178 of consensus"
repeat_region 38800..39708
/note="LIME1 repeat: matches 2682..3528 of consensus"
repeat_region 39709..40383
/note="LIR8 repeat: matches 1..691 of consensus"
repeat_region 40384..42758
/note="LIME1 repeat: matches 3528..6105 of consensus"
repeat_region 42845..42931
/note="MERSB repeat: matches 51..133 of consensus"
repeat_region 42960..43209
/note="AIUy repeat: matches 1..249 of consensus"
repeat_region 43210..43273
/note="MERSB repeat: matches 2..64 of consensus"
repeat_region 43585..43732
/note="FRAM repeat: matches 1..148 of consensus"
repeat_region 44578..44891
/note="AIUto repeat: matches 1..304 of consensus"
repeat_region 45144..45444
/note="AIUSx repeat: matches 1..299 of consensus"
repeat_region 45473..46079
/note="LIP10 repeat: matches 5546..6160 of consensus"
repeat_region 46406..46483

```

```

repeat_region /note="MERS4 repeat: matches 1..81 of consensus"
48101..48261
/note="MERS1B repeat: matches 2..158 of consensus"
50465..50568
/note="MIR repeat: matches 15..123 of consensus"
50987..51068
/note="MLTIE repeat: matches 50..133 of consensus"
51171..51429
/note="MLTIE repeat: matches 265..557 of consensus"
52002..52259
/note="AIUSx repeat: matches 33..290 of consensus"
52819..53599
/note="LIP2 repeat: matches 2..776 of consensus"
53595..56841
/note="LIP2 repeat: matches 900..6146 of consensus"
59147..59190

Query Match 1.5%: Score 20; DB 9; Length 70339;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1226 ctaataattgatctaaag 1245
Db 36036 CTAATAATTGATCTAAG 36055
|||||
|||||

RESULT 29
AC005694 82101 bp DNA PRI 01-JUN-2000
LOCUS Homo sapiens chromosome 22q12 clone bk294c2, complete sequence.
DEFINITION AC005694
ACCESSION AC005694.3 GI:4996908
VERSION HTG.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Unpublished
Homo sapiens Chromosome 22q11 BAC Clone bk294c2 in MDR Region
2 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (23-SEP-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (04-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (19-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
5 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (23-DEC-1998) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
6 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
Direct Submission
Submitted (05-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
7 (bases 1 to 82101)
Fang,F., Fu,Y., Pan,H. and Roe,B.A.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

TITLE Direct Submission
JOURNAL Submitted (10-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 8 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (12-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 9 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 10 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (26-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 11 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (28-JAN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 12 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (06-FEB-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 13 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (23-FEB-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 14 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (16-MAR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 15 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (18-MAR-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 16 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (04-JUN-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 17 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (24-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 18 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (26-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 19 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission

JOURNAL Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS 20 (bases 1 to 82101)
TITLE Fang,F., Fu,Y., Pan,H. and Roe,B.A.
JOURNAL Direct Submission
Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jun 4, 1999 this sequence version replaced gi:4263797.
Because these overlapping clones came from different libraries
there are numerous instances of insertions, deletions, and single
nucleotide polymorphisms in the overlapping regions below
AC005529(bx256d12) 96441 180736 (137732) overlaps AC005694(bx294c2)
1 82091 (10) AC005527(489d1) 98413 149308 (0) overlaps
AC005694(bx294c2) 1 50869 (31232) AC002991(n47n6) 26513 38140 (0)
overlaps AC005694(bx294c2) 1 11605 (70496).
location/Qualifiers
1..82101
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22q12"
/clone="BX294C2"
FEATURES
SOURCE
BASE COUNT 21863 a 18718 c 19364 g 22156 t.
ORIGIN
Query Match 1.5%; Score 20; DB 9; Length 82101;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 711 tgaattggtgcacagttga 730
|||||
Db 63545 TGAATGCTGCAGAGTTGA 63564
RESULT 30
AC007164
LOCUS
DEFINITION Homo sapiens BAC clone RP11-304A10 from 7p21-p22, complete
sequence.
ACCESSION AC007164
VERSION AC007164.3 GI:7243907
KEYWORDS HTG.
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 89862)
TITLE Sulston,J.E. and Waterston,R.
JOURNAL Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
AUTHORS 2 (bases 1 to 89862)
TITLE Holmes,A., Kallio,J., Pape,K. and Jones,T.
JOURNAL The sequence of Homo sapiens BAC clone RP11-304A10
unpublished
REFERENCE
AUTHORS 3 (bases 1 to 89862)
TITLE Waterston,R.H.
JOURNAL Direct Submission
Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 4 (bases 1 to 89862)
TITLE Waterston,R.H.
JOURNAL Direct Submission
Submitted (15-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 89862)
TITLE Waterston,R.
JOURNAL Direct Submission
Submitted (27-APR-2000) Department of Genetics, Washington

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 15, 2000 this sequence version replaced g1:4662667.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0304A10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-668E10, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-304A10. Actual end is at base position 62268 of RP4-668E10.

Location/Qualifiers

1. 89862

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7p21-p22"

/clone="RP11-304A10"

/clone_lib="RPI1-11"

340. .457

/rpt_family="MIR"

461. .512

/rpt_family="MIR"

513. .921

/rpt_family="MIR"

922. .1031

/rpt_family="MIR"

2360. .2834

/rpt_family="Retroviral"

2889. .3572

/rpt_family="L2"

3825. .3923

/rpt_family="MIR"

5967. .6042

/rpt_family="MIR"

6530. .6990

repeat_region

repeat_region /rpt_family="L1"
6998. .7186 /rpt_family="L1"
7240. .7318 /rpt_family="L1"
7319. .7391 /rpt_family="L1"
7392. .7749 /rpt_family="L1"
7750. .7945 /rpt_family="L1"
7905. .8499 /rpt_family="L1"
note="match to EST AW003673 (NTD:95850789) ws61d12.x1"
7966. .8085 /rpt_family="MIR"
8086. .8227 /rpt_family="MIR"
8431. .9531 /rpt_family="Retroviral"
9763. .9867 /rpt_family="MIR"
10886. .10907 /rpt_family="MIR"
11444. .11549 /rpt_family="AT-rich"
12509. .12539 /rpt_family="Alu"
12554. .13483 /rpt_family="TTTA)n"
14143. .14638 /rpt_family="L1"
14711. .14882 /rpt_family="L1"
15988. .16200 /rpt_family="MIR"
16325. .16627 /rpt_family="MIR"
18028. .18081 /rpt_family="Alu"
18135. .18384 /rpt_family="(TG)n"
19044. .19278 /rpt_family="MIR"
19721. .19800 /rpt_family="MIR"
19825. .20135 /rpt_family="CT-rich"
21951. .22034 /rpt_family="L2"
22071. .22364 /rpt_family="L2"
23130. .23251 /rpt_family="Alu"
24315. .24519 /rpt_family="L2"
25403. .25429 /rpt_family="MIR"
27193. .27234 /rpt_family="AT-rich"
27400. .29368 /rpt_family="AT-rich"
29376. .29535 /rpt_family="L1"
29536. .29626 /rpt_family="Alu"
29631. .29753 /rpt_family="(TA)n"
29935. .30283 /rpt_family="L1"
30306. .30385 /rpt_family="L1"
33392. .33417 /rpt_family="Alu"
33392. .33417 /rpt_family="AT-rich"

```

repeat_region      33435..33473
                    /rpt_family="L2"
repeat_region      33481..33655
                    /rpt_family="MER1_type"
repeat_region      33690..33988
                    /rpt_family="Alu"
repeat_region      34079..34402
                    /rpt_family="L1"
repeat_region      34416..34585
                    /rpt_family="(TA)n"
repeat_region      34614..34670
                    /rpt_family="(TA)n"
repeat_region      34956..35480
                    /rpt_family="L2"
repeat_region      35556..35687
                    /rpt_family="MER1_type?"
repeat_region      35742..35873
                    /rpt_family="MER1_group"
repeat_region      36352..36403

```

```

Query Match      1.5%, Score 20; DB 9; Length 89862;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 613 gctgatttaattgcttgc 632
Db 77206 GCTGATTTAATTGCTTGC 77225

```

```

RESULT 31
AL354692/2c      110443 bp      DNA      PRI      25-JUL-2001
LOCUS      Human DNA sequence from clone RP11-361M4 on chromosome 9, complete
DEFINITION      sequence.
ACCESSION      AL354692
VERSION      AL354692.23 GI:15022181
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 110443)
AUTHORS      Wilson,S.
TITLE      Direct Submision
JOURNAL      Submitted (25-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 26, 2001 this sequence version replaced gi:15021138.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone compis of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/Hgp/Cnr9
RP11-361M4 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

```

VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-361M4. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-361M4 is at 108444 in this
 sequence. The true right end of clone RP11-59K22 is at 2000 in this
 sequence.

```

FEATURES
source      1..110443
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="9"
            /clone="RP11-361M4"
            /clone_11b="RPCI-11.2"
repeat_region      1..4068
                    /note="LIP2 repeat: matches 1478..5546 of consensus"
repeat_region      4069..4362
                    /note="AluX repeat: matches 1..293 of consensus"
repeat_region      4363..5588
                    /note="LIP2 repeat: matches 258..1478 of consensus"
repeat_region      5566..5949
                    /note="LIM1 repeat: matches 2018..2400 of consensus"
repeat_region      5950..6235
                    /note="AluX repeat: matches 1..303 of consensus"
repeat_region      6236..9260
                    /note="LIM1 repeat: matches -1390..2018 of consensus"
repeat_region      9684..9810
                    /note="LIM4 repeat: matches 5317..5431 of consensus"
repeat_region      9835..10361
                    /note="LIP16 repeat: matches 5636..6150 of consensus"
repeat_region      10362..10669
                    /note="AluX repeat: matches 1..309 of consensus"
repeat_region      10670..11196
                    /note="LIP16 repeat: matches 5084..5636 of consensus"
repeat_region      12642..12750
                    /note="MIR repeat: matches 54..168 of consensus"
repeat_region      12824..13019
                    /note="L1 repeat: matches 2639..2830 of consensus"
repeat_region      13195..13426
                    /note="116 copies 2 mer ta 63% conserved"
repeat_region      13449..13617
                    /note="L1 repeat: matches 2296..2464 of consensus"
repeat_region      14554..14854
                    /note="AluX repeat: matches 1..303 of consensus"
repeat_region      15520..15667
                    /note="MER63A repeat: matches 25..172 of consensus"
repeat_region      15863..16146
                    /note="AluB repeat: matches 1..286 of consensus"
repeat_region      16205..16533
                    /note="AluX repeat: matches 1..296 of consensus"
repeat_region      18551..18875
                    /note="MER69 repeat: matches 1..336 of consensus"
repeat_region      18903..19120
                    /note="MER69 repeat: matches 405..629 of consensus"
repeat_region      19826..20125
                    /note="AluX repeat: matches 1..301 of consensus"
repeat_region      20374..20752
                    /note="LIT1J repeat: matches 5..369 of consensus"
repeat_region      20859..22019
                    /note="LIMB repeat: matches 114..1295 of consensus"
repeat_region      22022..22088
                    /note="Alu repeat: matches 230..293 of consensus"
repeat_region      22092..22257
                    /note="L1 repeat: matches 4789..4975 of consensus"
repeat_region      26196..26223
                    /note="14 copies 2 mer tg 100% conserved"
repeat_region      27712..27915
                    /note="LIM4 repeat: matches 6004..6211 of consensus"
repeat_region      29575..29898
                    /note="MER33 repeat: matches 2..324 of consensus"
repeat_region      30358..30404
                    /note="MIR repeat: matches 117..163 of consensus"
repeat_region      30500..30567

```



```

repeat_region /note="MIR repeat: matches 64. .135 of consensus"
32058. .32383
repeat_region /note="MER33 repeat: matches 2. .324 of consensus"
34237. .34337
repeat_region /note="MIR repeat: matches 20. .123 of consensus"
34386. .34553
repeat_region /note="PRAM repeat: matches 4. .172 of consensus"
35130. .35249
repeat_region /note="MIR repeat: matches 15. .136 of consensus"
35278. .35408
repeat_region /note="MIR repeat: matches 124. .258 of consensus"
35513. .35591
repeat_region /note="L2 repeat: matches 2630. .2708 of consensus"
36246. .36463
repeat_region /note="MER58A repeat: matches 5. .224 of consensus"
36474. .36567
repeat_region /note="MIR repeat: matches 89. .190 of consensus"
38006. .38530
repeat_region /note="MLTID repeat: matches 9. .505 of consensus"
39325. .39622
repeat_region /note="Alusg repeat: matches 1. .239 of consensus"
41143. .41277
repeat_region /note="MIR repeat: matches 92. .227 of consensus"
41278. .41380
repeat_region /note="MLTID repeat: matches 402. .505 of consensus"
41417. .41608
repeat_region /note="MLTID repeat: matches 13. .205 of consensus"
42905. .43159
repeat_region /note="LTRIC repeat: matches 114. .387 of consensus"
44346. .44687
repeat_region /note="L2 repeat: matches 2361. .2697 of consensus"
44688. .44830
repeat_region /note="MLTIE repeat: matches 398. .567 of consensus"
46666. .45880
repeat_region /note="LIM4 repeat: matches 3128. .3374 of consensus"
45928. .45979
repeat_region /note="LIM4 repeat: matches 2958. .3011 of consensus"
46147. .46446
repeat_region /note="Alusx repeat: matches 1. .299 of consensus"
46929. .47027
repeat_region /note="L2 repeat: matches 2607. .2706 of consensus"
47128. .47308
repeat_region /note="L2 repeat: matches 2290. .2475 of consensus"
47379. .47555
repeat_region /note="LIM4c repeat: matches 1245. .1416 of consensus"
47584. .48131
repeat_region /note="LIM4c repeat: matches 1499. .2065 of consensus"
48784. .49492
repeat_region /note="LIME repeat: matches 2407. .3139 of consensus"
49581. .49870
repeat_region /note="LIME repeat: matches 3292. .3573 of consensus"
49871. .49985
repeat_region /note="LIMC3 repeat: matches 7623. .7740 of consensus"
49986. .51673
repeat_region /note="LIME repeat: matches 3573. .5320 of consensus"
51751. .51944
repeat_region /note="LIME repeat: matches 5509. .5707 of consensus"
53856. .54072
repeat_region /note="MER20 repeat: matches 1. .218 of consensus"
55390. .55802
repeat_region /note="LTR33 repeat: matches 38. .435 of consensus"
57393. .57693
repeat_region /note="Alusp repeat: matches 1. .302 of consensus"
63187. .63299
repeat_region /note="L2 repeat: matches 2632. .2750 of consensus"
63321. .64037
repeat_region /note="L2 repeat: matches 1913. .2710 of consensus"
65180. .66203
repeat_region /note="L2 repeat: matches 240. .1234 of consensus"
66204. .66477
repeat_region /note="MLTIA1 repeat: matches 1. .365 of consensus"
66550. .67584
repeat_region /note="L2 repeat: matches 1431. .2481 of consensus"

```

```

repeat_region 67581. .67651
repeat_region /note="MIR repeat: matches 142. .212 of consensus"
69580. .69887
repeat_region /note="Aluy repeat: matches 1. .307 of consensus"
71133. .71329
repeat_region /note="MIR repeat: matches 55. .247 of consensus"
71439. .71603
repeat_region /note="MIR repeat: matches 5. .181 of consensus"
71633. .71686
repeat_region /note="L2 repeat: matches 2646. .2703 of consensus"
71960. .72259
repeat_region /note="Aludb repeat: matches 1. .312 of consensus"

Query Match 1.5%; Score 20; DB 9; Length 110443;
Best Local Similarity 100.0%; Prid. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 61560 CTAATTAATTGATCTCTAG 61547
Oy 1226 ctaataattgatctcctaag 1245
|||||
|||||

RESULT 32
AC011285
LOCUS
DEFINITION Homo sapiens chromosome CTD clone CTD-2337B17, WORKING DRAFT
ACCESSION AC011285.5 GI:9954810
VERSION AC011285.5
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 120891)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 120891)
AUTHORS Waterston,R.H.
TITLE Direct Submission
COMMENT Submitted (05-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 31, 2000 this sequence version replaced g1:9838031.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MS2337B17
----- Summary Statistics -----
Sequencing vector: M13, 92%
Sequencing vector: Plasmid: 8%
Chemistry: Dye-primer ET: 2% of reads
Chemistry: Dye-terminator Big Dye: 8% of reads
Assembly program: Phrap; version 0.99013
Consensus quality: 116247 bases at least Q40
Consensus quality: 120875 bases at least Q30
Consensus quality: 123092 bases at least Q20
Insert size: 92000; agarose-efp
Insert size: 120440; sum-of-configs
Quality coverage: 15.35 in Q20 bases; sum-of-configs
Quality coverage: 15.06 in Q20 bases; sum-of-configs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```



```

misc_feature      /note="assembly_name:Contig122"
                  21093..22827
                  /note="assembly_name:Contig123"
misc_feature      /note="assembly_name:Contig123"
                  22928..24125
                  /note="assembly_name:Contig124"
misc_feature      /note="assembly_name:Contig124"
                  24226..27014
                  /note="assembly_name:Contig125"
misc_feature      /note="assembly_name:Contig125"
                  27115..28847
                  /note="assembly_name:Contig126"
misc_feature      /note="assembly_name:Contig126"
                  28948..30686
                  /note="assembly_name:Contig127"
misc_feature      /note="assembly_name:Contig127"
                  30787..3120891
                  /note="assembly_name:Contig129
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      39086 a 22644 c 22727 g 34516 t 1918 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 120891;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 7 atattcaatgttatttta 26
    ||||||||||||||||
Db 51929 ATATCAAAAGTTATTTA 51948

RESULT 33
LOCUS      HSA118000 126138 bp DNA PRT 05-MAY-1999
DEFINITION Homo sapiens NF2 gene.
ACCESSION Y18000
VERSION Y18000.1 GI:3980299
KEYWORDS  Alu-like repetitive sequence; L1 repeat; Line repeat; MER repeat;
SOURCE      MIR repeat; NF2 gene; repetitive sequence.
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 126138)
Zucman-Rossi,J., Legoux,P., Der Sarkissian,H., Cheret,G., Sor,F.,
Bernardi,A., Cazes,L., Giraud,S., Ollagnon,E., Lenoir,G. and
Thomas,G.
NF2 gene in neurofibromatosis type 2 patients
Hum. Mol. Genet. 7 (13), 2095-2101 (1998)
99035680
2 (bases 1 to 126138)
Legoux,P., Legrand,M.F., Ollagnon,E., Lenoir,G. and
Zucman-Rossi,J.
Characterisation of 16 polymorphic markers in the NF2 gene:
application to hemizyosity detection
Hum. Mutat. 13 (4), 290-293 (1999)
99235548
3 (bases 1 to 126138)
Zucman-Rossi,J.
Direct Submission
Submitted (25-AUG-1998) J. Zucman-Rossi, INSERM U434/CEPH, 27 rue
Ulfette Docu, 75010 Paris, FRANCE
location/Qualifiers
    1..126138
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    31..171
    /rpt_family="LMD2"
    /rpt_unit=31..171
    /rpt_type=DISPERSED
    206..486
    /rpt_family="Alu-S2"
    /rpt_unit=206..486
    /rpt_type=DISPERSED
    881..1161
    /rpt_family="Alu-Sg"

```

```

/rpt_unit=884..1161
/rpt_type=DISPERSED
repeat_region 1239..1526
/rpt_family="Alu-Sg"
/rpt_unit=1239..1526
/rpt_type=DISPERSED
repeat_region 1564..1848
/rpt_family="Alu-Jo"
/rpt_unit=1564..1848
/rpt_type=DISPERSED
repeat_region 1868..1922
/rpt_family="MUTIE"
/rpt_unit=1868..1922
/rpt_type=DISPERSED
repeat_region 1923..2079
/rpt_family="Alu-Jb"
/rpt_unit=1923..2079
/rpt_type=DISPERSED
LTR 2124..2809
/note="LTR 12"
LTR 2811..2974
/note="LTR 12"
repeat_region 2975..3565
/rpt_family="PER5"
/rpt_unit=2975..3565
/rpt_type=DISPERSED
repeat_region 3574..3704
/rpt_family="Alu-Jb"
/rpt_unit=3574..3704
/rpt_type=DISPERSED
repeat_region 3730..3935
/rpt_family="MER20"
/rpt_unit=3730..3935
/rpt_type=DISPERSED
repeat_region 3981..4117
/rpt_family="MIR"
/rpt_unit=3981..4117
/rpt_type=DISPERSED
repeat_region 4120..4403
/rpt_family="Alu-Sx"
/rpt_unit=4120..4403
/rpt_type=DISPERSED
repeat_region 4425..4489
/rpt_family="MIR"
/rpt_unit=4425..4489
/rpt_type=DISPERSED
repeat_region 4772..4864
/rpt_family="Alu-Sg"
/rpt_unit=4772..4864
/rpt_type=DISPERSED
repeat_region 4969..4988
/rpt_family="Alu-Sx"
/rpt_unit=4969..4988
/rpt_type=DISPERSED
repeat_region 4988..5276
/rpt_family="Alu-Y"
/rpt_unit=4988..5276
/rpt_type=DISPERSED
repeat_region 5317..5463
/rpt_family="Alu-Jo"
/rpt_unit=5317..5463
/rpt_type=DISPERSED
repeat_region 5470..5760
/rpt_family="Alu-Sz"
/rpt_unit=5470..5760
/rpt_type=DISPERSED
repeat_region 5781..5908
/rpt_family="Alu-Jo"
/rpt_unit=5781..5908
/rpt_type=DISPERSED
repeat_region 6197..6479
/rpt_family="Alu-Y"
/rpt_unit=6197..6479
/rpt_type=DISPERSED
repeat_region 6869..7036
/rpt_family="MIR"
/rpt_unit=6869..7036
/rpt_type=DISPERSED
repeat_region 7063..7351
/rpt_family="Alu-Sg"
/rpt_unit=7063..7351
/rpt_type=DISPERSED
repeat_region 7668..7949
/rpt_family="Alu-Sp"
/rpt_unit=7668..7949
/rpt_type=DISPERSED
repeat_region 7950..7990
/rpt_family="trinucleotide repeat, taa"
/rpt_unit=7950..7990
/rpt_type=DISPERSED
variation 8240
/note="polymorphism"
/frequency=".58"
/replace="C"
variation 8348
/note="polymorphism"
/frequency=".54"
/replace="tttttt"
variation 8787
/note="polymorphism"
/frequency=".43"
/replace="a"
exon 8819..9104
/gene="NF2"
/number=1
8819..99278
/gene="NF2"
variation 8881
/gene="NF2"
/note="polymorphism"
/frequency=".31"
/replace="g"
CDS 59142..59210,60079..60161,62673..62748,65689..65823,
69470..69544,72814..72927,76297..76419,77740..77957,
79307..79412,82675..82802,85918..86080,99228..99278)
/gene="NF2"
/note="isoform 1"
/codon_start=1
/product="NF2 protein"
/protein_id="CA76992.1"
/db_xref="GI:3980300"
/translation="MAGAIASRMSESSLRKRPKPTVRIIVTMDAEMEFNCMKRKG
DLPLVORTGLRETWFEGLOYTKDVIAMKMKKVLHDVSKSEPVTHPLAKFP
ENABEELVOZITQHLFLOYKKQITLDEKICPPASVILASVAQAKGYDPSVHR
GFLAOEELPKRVINLTOMPEMEERITAHYHRGRADDEAMEILTKAOLEMKG
DKIDVFNKSSKLRVNLILQICIGNHLMRRKADSLEVOQMKQAQAEARKOM
EORLAREKOMEAEERTRDELRLRLHOMKEAAMAEALMRSETDALLAQAITE
EERKLAKAALAEQEMORIKATAIRTEEEKRLQCEVLEVALMAAESERRAKE
ADOLKODLQAEARBAERKOKLLEIATKPTYPNPIAPLPIDIPFNIGDSISD
EKPDMDKLSMEIRKEVEYEMKSKHLOEOLNEKLEIEMALKERETALDILHENS
DRGSSKNTTKTKTLDSAKSRVAFPEEL"
59142..59210,60079..60161,62673..62748,65689..65823,
69470..69544,72814..72927,76297..76419,77740..77957,
79307..79412,82675..82802,85918..86080,87501..87611)
/gene="NF2"
/note="isoform II"
/codon_start=1
/product="NF2 protein"
/protein_id="CA76993.1"
/db_xref="GI:3980301"
/translation="MAGAIASRMSESSLRKRPKPTVRIIVTMDAEMEFNCMKRKG
DLFDLCRTGLRETWFEGLOYTKDVIAMKMKKVLHDVSKSEPVTHPLAKFP
ENABEELVOZITQHLFLOYKKQITLDEKICPPASVILASVAQAKGYDPSVHR
GFLAOEELPKRVINLTOMPEMEERITAHYHRGRADDEAMEILTKAOLEMKG
VNFPIRNKKGTGLLDALGLHIYDENLITRKISPPWNEINISISDSEFIKPL

```

```

DKTIDVERNSKLRVKNKILICIGNHDLFMRRKADSLFEOOMKQAREKAROM
EROLARERKREAREKREKREKREKREKREKREKREKREKREKREKREKREK
EAKLLAKRAAEQEMOKRATIRTEBEKRLMEQVLAELVATLMAESERRAE
ADJLKQDLQKAREKRAKOLLEIATKPPYPPNPPIPALPPIPSFNLDLSLSD
FKDDMKRLSMEIKEREYMEKRLQEODELKTEIEALKLERETADILHNENS
DRGSSKHNITIKSSPROKTYLHLSPOSRLFPGLTYVMVLPVILTRAR"
9105..41238
/number="1
/gene="NF2"
repeat_region
9779..9912
/rpl_family="MER88A"

Query Match      1.5%; Score 20; DB 9; Length 126138;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 711 tgaattggtcagagattga 730
|||||
Db 38780 TGAATTGTCAGAGATTGA 38799

RESULT 34
AC007559/c 156374 bp DNA PRI 30-SEP-2000
LOCUS Homo sapiens BAC clone RP11-364J6 from 2, complete sequence.
DEFINITION AC007559
AC007559.3 GI:6671987
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 156374)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 156374)
AUTHORS Kyung, K., Stoneking, T., Nguyen, C. and Hawrysko, C.
TITLE The sequence of Homo sapiens BAC clone RP11-364J6
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 156374)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 156374)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 156374)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 5, 2000 this sequence version replaced gi:5732153.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.edu
----- Summary Statistics
-----
Center project name: H.NH0364J06

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

FEATURES

source

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frenzen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-15002, 200 bp overlap. Actual start of this clone is at base position 119635 of RP11-15002; actual end is at base position 156374 of RP11-364J6. Location/Qualifiers

```

1..156374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-364J6"
/clone_id="RP11-11"
859..982
/rpl_family="MER94"
1101..1234
/rpl_family="Alu"
1235..1541
/rpl_family="Alu"
1542..1674
/rpl_family="Alu"
1675..1817
/rpl_family="(GAGAA)n"
2425..2856
/note="match to EST AA457626 (NID:92180346) aa89e10.r1"
misc-feature
repeat_region
/rpl_family="MER1-type"
4452..4500
/rpl_family="GA-rich"
4626..5212
/rpl_family="L1"
5383..5663
/rpl_family="L2"
5683..5804
/rpl_family="L2"
5839..6177
/rpl_family="L2"
repeat_region
8592..8703
/rpl_family="MALR"
8770..8994
/rpl_family="MALR"
9506..9681
/rpl_family="MIR"
9682..9973
/rpl_family="(CAT)n"
9990..10112
/rpl_family="(CAT)n"
10210..10389
/rpl_family="(CAT)n"
10404..10580
repeat_region

```

```

repeat_region      /rpt_family="(CAT)n"
10598..11147
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11159..11338
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11350..11572
/rpt_family="(CAT)n"
repeat_region      /rpt_family="(CAT)n"
11514..11954
/rpt_family="(CAT)n"
misc_feature        /rpt_family="(CAT)n"
11558..12135
/note="match to EST AA425721 (NTD:92106441) zw49b10.s1"
13250..13597
/note="match to EST AA427791 (NTD:92111607) zw49b10.r1"
repeat_region      /rpt_family="(L2)"
13735..14214
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15036..15174
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15560..15686
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15777..15960
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
15981..16065
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
16066..16366
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
16367..16539
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17008..17175
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17176..17463
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
17464..18111
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
19413..19447
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
19766..19856
/rpt_family="(L2)"
misc_feature        /rpt_family="(L2)"
21086..21424
/note="match to EST AA494044 (NTD:92233885) ng60e10.s1"
21133..21172
/rpt_family="(L2)"
misc_feature        /rpt_family="(L2)"
21430..21451
/note="match to EST AA494044 (NTD:92233885) ng60e10.s1"
21595..21634
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
21911..21954
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
24210..24476
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
26241..26264
/rpt_family="(L2)"
misc_feature        /rpt_family="(L2)"
26751..26970
/note="match to EST AA476160 (NTD:94329205) tn62h08.x1"
28700..28725
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
29474..29763
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
30021..30386
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31035..31121
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31494..31719
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
31720..31758
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
32071..32178
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
32804..32511
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
33334..33375
/rpt_family="(L2)"
repeat_region      /rpt_family="(L2)"
33422..33441
/rpt_family="(L2)"

```

```

repeat_region      33679..33739
/rpt_family="(L2)"
repeat_region      34762..34820
/rpt_family="(L2)"
Query Match        1.5%; Score 20; DB 9; Length 156374;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1181 gagctgcagaattcaagttg 1200
Db 52151 GAGCTGCAGATTCAAGTTG 52132

```

```

RESULT 35
AL596283
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

COMMENT
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Project Information
Center project name: h0257N2

```

```

----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 158737 bases at least Q40
Consensus quality: 159906 bases at least Q30
Consensus quality: 160628 bases at least Q20
Insert size: 161323; sum-of-contigs
Insert size: 165641; 3.6% error; agarose-tp
Quality coverage: 6.59x in Q20 bases; sum-of-contigs quality
coverage: 6.45x in Q20 bases; agarose-tp
-----

```

```

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

FEATURES
Source
Location/Qualifiers
1..163323
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="XXbac-257N2"
/clone_lib="NOD mouse library"
1..6428
/note="assembly-fragment:01729"
fragment_chain:1
6529..14784
/note="assembly-fragment:02816"
fragment_chain:1
14885..37995
/note="assembly-fragment:01187.0"
38096..55258
/note="assembly-fragment:00197"
fragment_chain:2

```

```

misc_feature
1..6428
/note="assembly-fragment:01729"
fragment_chain:1
6529..14784
/note="assembly-fragment:02816"
fragment_chain:1
14885..37995
/note="assembly-fragment:01187.0"
38096..55258
/note="assembly-fragment:00197"
fragment_chain:2

```

```

misc_feature
1..6428
/note="assembly-fragment:01729"
fragment_chain:1
6529..14784
/note="assembly-fragment:02816"
fragment_chain:1
14885..37995
/note="assembly-fragment:01187.0"
38096..55258
/note="assembly-fragment:00197"
fragment_chain:2

```

```

misc_feature
1..6428
/note="assembly-fragment:01729"
fragment_chain:1
6529..14784
/note="assembly-fragment:02816"
fragment_chain:1
14885..37995
/note="assembly-fragment:01187.0"
38096..55258
/note="assembly-fragment:00197"
fragment_chain:2

```

```

misc_feature      52629..71295
                  /note="assembly_fragment:01892
                  fragment_chain:2"
misc_feature      71396..78245
                  /note="assembly_fragment:02371
                  fragment_chain:2"
misc_feature      78346..103280
                  /note="assembly_fragment:01370
                  fragment_chain:2"
misc_feature      103381..135684
                  /note="assembly_fragment:01702
                  fragment_chain:2"
misc_feature      135785..146174
                  /note="assembly_fragment:02231
                  fragment_chain:2"
misc_feature      146275..160087
                  /note="assembly_fragment:03006
                  fragment_chain:2"
misc_feature      160188..162323
                  /note="assembly_fragment:01168
                  fragment_chain:2
                  clone_end:SP6
                  vector_side:right"
BASE COUNT      45002 a 33182 c 33907 g 49227 t 1005 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 162323;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      329 gtgacattatcagactcta 348
        |||||||||||||||||||
Db 128540 GTGACATTATTCAGACTCTA 128559

RESULT 36
AC073556/c      163189 bp      DNA      HTG      16-AUG-2001
LOCUS
DEFINITION      Oryza sativa chromosome 3 clone OSJNBa0091P11. *** SEQUENCING IN
PROGRESS ***. 1 ordered pieces.
ACCESSION      AC073556
VERSION      AC073556.8
KEYWORDS      HTG; HTGS_PHASE2.
SOURCE      Oryza sativa.
ORGANISM      Oryza sativa
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE      1 (bases 1 to 163189)
               Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
               Gansberger, K., Burgess, S., Jarrahi, B., Shwartsbeyn, M., Brenner, M.,
               Clacko, A., Pal, G., Vanaken, S., Hansen, C., Uteirbach, T.,
               Feldblyum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
               Salzberg, S. and Fraser, C.
               Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0091P11 BAC genomic
               sequence
               Unpublished
               2 (bases 1 to 163189)
               Buell, R.
               Direct Submission
               Submitted (23-JUN-2000) The Institute for Genomic Research, 9712
               Medical Center Dr., Rockville, MD 20850, USA
               On Aug 16, 2001 this sequence version replaced gi:14993751.
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 1 contigs. Gaps between the contigs
               * are represented as runs of N. The order of the pieces
               * is believed to be correct as given, however the sizes
               * of the gaps between them are based on estimates that have
               * provided by the submitter.
               * This sequence will be replaced
               * by the finished sequence as soon as it is available and
               * the accession number will be preserved.

```

```

FEATURES
Source
*      1 163189: contig of 163189 bp in length.
Location/Qualifiers
1..163189
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="3"
/clone="OSJNBa0091P11"
BASE COUNT      47068 a 35531 c 35052 g 45534 t 4 others
ORIGIN
Query Match      1.5%; Score 20; DB 2; Length 163189;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      905 gtgctatttcttgaagt 924
        |||||||||||||||||||
Db 73824 GTGCTATTCTTGTGAAGT 73805

```

```

RESULT 37
AC016845/c      170294 bp      DNA      HTG      11-AUG-2001
LOCUS
DEFINITION      Homo sapiens clone RP11-5E23. *** SEQUENCING IN PROGRESS ***. 5
unordered pieces.
ACCESSION      AC016845
VERSION      AC016845.6
KEYWORDS      HTG; HTGS_PHASE1; HTGS_FULFUP; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 170294)
               Birren, B., Linton, L., Nusbaum, C. and Lander, E.
               Unpublished
               2 (bases 1 to 170294)
               Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
               Baldridge, J., Barina, N., Beckert, R., Boguslavsky, L., Boukhalter, B.,
               Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
               Cooke, P., DeRellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
               Ferreira, P., Fitzhugh, W., Forrest, C., Funke, K., Gage, D.,
               Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
               Howland, J., Johnson, R., Jones, C., Kane, L., Karatas, A., Klein, J.,
               Lechoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
               McEwan, P., McGurk, A., McKernan, K., McDonald, J., Meldrum, J.,
               Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
               Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
               Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
               Tesitaye, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
               Wyman, D., Ye, W., Zimmer, A. and Zody, M.
               Direct Submission
               Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               On Aug 11, 2001 this sequence version replaced gi:11990706.
               All repeats were identified using RepeatMasker:
               http://ftp.genome.washington.edu/RV/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: MIBR
               Web site: http://www.seg.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: 5_E_23
               Center clone name: 5_E_23
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 5 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as

```



```

STS
repeat_region /rpt_family="L1" complement(5091..5707)
                /db_xref="GI:6126790"
repeat_region 5329..5549 /rpt_family="L1"
repeat_region 5578..6135 /rpt_family="L1"
repeat_region 6230..6913 /rpt_family="L1"
repeat_region 6914..7288 /rpt_family="Retroviral"
repeat_region 7365..7613 /rpt_family="L1"
repeat_region 7614..7916 /rpt_family="L1"
repeat_region 7917..7948 /rpt_family="Alu"
repeat_region 7949..8246 /rpt_family="L1"
repeat_region 8247..8510 /rpt_family="L1"
repeat_region 8640..8933 /rpt_family="Alu"
repeat_region 8937..9246 /rpt_family="Alu"
repeat_region 9247..9276 /rpt_family="Alu"
repeat_region 9567..10019 /rpt_family="(TAAA)n"
repeat_region 10220..10387 /rpt_family="MALR"
repeat_region 10388..10433 /rpt_family="Alu"
repeat_region 10434..10713 /rpt_family="(TTTA)n"
repeat_region 10714..10840 /rpt_family="Alu"
repeat_region 10843..11060 /rpt_family="L1"
repeat_region 11325..11357 /rpt_family="L1"
repeat_region 11358..11641 /rpt_family="(TTTA)n"
repeat_region 11823..11916 /rpt_family="Alu"
repeat_region 11917..11937 /rpt_family="L1"
repeat_region 11938..12224 /rpt_family="(TTTC)n"
repeat_region 12225..12361 /rpt_family="Alu"
repeat_region 12394..12526 /rpt_family="L1"
repeat_region 12527..12816 /rpt_family="Alu"
repeat_region 12817..12933 /rpt_family="Alu"
repeat_region 13066..13088 /rpt_family="Alu"
repeat_region 13089..13237 /rpt_family="(TTTC)n"
repeat_region 13402..13427 /rpt_family="Alu"
repeat_region 13433..13681 /rpt_family="(T)n"
repeat_region 13841..14021 /rpt_family="Alu"
repeat_region 14045..14178 /rpt_family="MERL_type"
repeat_region 14221..14269 /rpt_family="L1"
repeat_region 14337..14616 /rpt_family="Alu"
                /rpt_family="Alu"

repeat_region 15001..15089 /rpt_family="L1"
repeat_region 15090..15411 /rpt_family="L1"
repeat_region 15412..15817 /rpt_family="Alu"
repeat_region 15944..16375 /rpt_family="L1"
repeat_region 16419..16482 /rpt_family="MALR"
repeat_region 16537..16565 /rpt_family="MIR"
repeat_region 16571..16801 /rpt_family="AT-rich"
repeat_region 16884..17128 /rpt_family="AT-rich"

Query Match 1.5% Score 20; DB 9; Length 173893;
Best Local Similarity 100.0%; Pred No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 784 ttgaaagattcacttacc 803
Db 45927 TTGGAAGATTCACTTACCA 45946

RESULT 39
AC022812
LOCUS AC022812 174596 bp DNA HTG 12-MAR-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-745f13 map 6, WORKING DRAFT
SEQUENCE 24 unordered pieces.
ACCESSION AC022812
VERSION AC022812.2 GI:7229851
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 174596)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 6, clone RP11-745f13
2 (bases 1 to 174596)
Unpublished
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Teisaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6922250.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6096

```

```

----- Summary Statistics -----
Center clone name: 745_F_13
Sequencing vector: M13: M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 161291 bases at least Q40
Consensus quality: 168261 bases at least Q30
Consensus quality: 170761 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 172296; sum-of-ctrls
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1284: contig of 1284 bp in length
* 1285 1384: gap of 100 bp
* 1385 2706: contig of 1322 bp in length
* 2707 2806: gap of 100 bp
* 2807 4528: contig of 1722 bp in length
* 4529 4628: gap of 100 bp
* 4629 6469: contig of 1841 bp in length
* 6470 6569: gap of 100 bp
* 6570 10731: contig of 4162 bp in length
* 10732 10831: gap of 100 bp
* 10832 13405: contig of 2574 bp in length
* 13406 13505: gap of 100 bp
* 13506 16970: contig of 3465 bp in length
* 16971 17070: gap of 100 bp
* 17071 20522: contig of 3452 bp in length
* 20523 20622: gap of 100 bp
* 20623 23195: contig of 2573 bp in length
* 23196 23295: gap of 100 bp
* 23296 27465: contig of 4170 bp in length
* 27466 27565: gap of 100 bp
* 27566 32567: contig of 5002 bp in length
* 32568 32667: gap of 100 bp
* 32668 37897: contig of 5230 bp in length
* 37898 37997: gap of 100 bp
* 37998 45060: contig of 7063 bp in length
* 45061 45160: gap of 100 bp
* 45161 50594: contig of 5433 bp in length
* 50595 50694: gap of 100 bp
* 50695 58198: contig of 7504 bp in length
* 58199 58298: gap of 100 bp
* 58299 68303: contig of 10005 bp in length
* 68304 68403: gap of 100 bp
* 68404 77724: contig of 9321 bp in length
* 77725 77824: gap of 100 bp
* 77825 86591: contig of 8767 bp in length
* 86592 86691: gap of 100 bp
* 86692 95521: contig of 8830 bp in length
* 95522 95621: gap of 100 bp
* 95622 105952: contig of 10331 bp in length
* 105953 106052: gap of 100 bp
* 106053 117439: contig of 11387 bp in length
* 117440 117539: gap of 100 bp
* 117540 132556: contig of 15017 bp in length
* 132557 132656: gap of 100 bp
* 132657 148474: contig of 15818 bp in length
* 148475 148574: gap of 100 bp
* 148575 174596: contig of 26022 bp in length.

```

```

FEATURES
SOURCE
1. 174596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"

```

```

/misc-feature /map="6"
/misc-feature /clone="RP11-745F13"
/misc-feature /clone_1b="RP11 Human Male BAC"
1. 1284
/misc-feature /note="assembly-fragment"
1385 2706
/misc-feature /note="assembly-fragment"
2807 4528
/misc-feature /note="assembly-fragment"
4629 6469
/misc-feature /note="assembly-fragment"
6570 10731
/misc-feature /note="assembly-fragment"
10832 13405
/misc-feature /note="assembly-fragment"
13506 16970
/misc-feature /note="assembly-fragment"
17071 20522
/misc-feature /note="assembly-fragment"
20623 23195
/misc-feature /note="assembly-fragment"
clone_end:17
vector_side:left"
23296 27465
/misc-feature /note="assembly-fragment"
27566 32567
/misc-feature /note="assembly-fragment"
32668 37897
/misc-feature /note="assembly-fragment"
37998 45060
/misc-feature /note="assembly-fragment"
45161 50594
/misc-feature /note="assembly-fragment"
clone_end:SP6
vector_side:left"
50695 58198
/misc-feature /note="assembly-fragment"
58299 68303
/misc-feature /note="assembly-fragment"
68404 77724
/misc-feature /note="assembly-fragment"
77825 86591
/misc-feature /note="assembly-fragment"
86692 95521
/misc-feature /note="assembly-fragment"
95622 105952
/misc-feature /note="assembly-fragment"
106053 117439
/misc-feature /note="assembly-fragment"
117540 132556
/misc-feature /note="assembly-fragment"
132657 148474
/misc-feature /note="assembly-fragment"
148575 174596
/misc-feature /note="assembly-fragment"

```

```

BASE COUNT 54131 a 33318 c 33201 g 51641 t 2305 others
ORIGIN

```

```

Query Match 1.5%; Score 20; DB 2: Length 174596;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1226 ctaataattgattcctaag 1245
|||||
DB 57509 CTAATAATTGATTCTAAG 57528

```

```

RESULT 40
AC007486 174725 bp DNA PRI 05-MAY-1999
LOCUS AC007486/c
DEFINITION Homo sapiens clone D1015P16A, complete sequence.
ACCESSION AC007486

```



```

VERSION AC007486.1 GI:4753270
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 174725)
JOURNAL The sequence of Homo sapiens clone
REFERENCE Waterston, R.H.
AUTHORS Unpublished
TITLE 2 (bases 1 to 174725)
JOURNAL Waterston, R.H.
AUTHORS Direct Submission
TITLE Submitted (05-MAY-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
FEATURES
Source
1..174725
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1015P16A"
BASE COUNT 58080 a 34168 c 31762 g 50715 t
ORIGIN
Query Match 1.5%; Score 20; DB 9; Length 174725;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 taticaatgtattttaa 27
|||||
Db 162503 TATTCAAATGTTATTTRAA 162484

RESULT 41
AC013261 175092 bp DNA HTG 30-OCT-2000
LOCUS Homo sapiens clone RP11-11E18, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION pieces.
AC013261
AC013261
AC013261.3 GI:11038541
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 175092)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE Homo sapiens, clone RP11-11E18
AUTHORS Unpublished
TITLE 2 (bases 1 to 175092)
JOURNAL Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Headford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McKernan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testafaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (05-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 30, 2000 this sequence version replaced g1:7107901.
All repeats were identified using RepeatMasker:
Smit, A.F., A. Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3142
Center clone name: L3142
----- Summary Statistics
Sequencing vector: M13; N77815; 48% of reads
Sequencing vector: plasmid; n/a; 52% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 174862 bases at least Q40
Consensus quality: 174949 bases at least Q20
Consensus quality: 174968 bases at least Q20
Insert size: 176000; agarose-tp
Insert size: 174992; sum-of-contigs
Quality coverage: 10.9 in Q20 bases; agarose-tp
Quality coverage: 11.0 in Q20
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86418: contig of 86418 bp in length
* 86419 86518: gap of 100 bp
* 86519 175092: contig of 88574 bp in length.
Location/Qualifiers
1..175092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11E18"
/clone_lib="RPCT-11 Human Male BAC"
1..86418
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
86519..175092
/note="assembly-fragment"
clone_end:T7
vector_side:right"
BASE COUNT 50202 a 34212 c 35769 g 54809 t 100 others
ORIGIN
Query Match 1.5%; Score 20; DB 2; Length 175092;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 ctataaatgtattcraag 1245
|||||
Db 39221 CTATTAATGATTCTTAAG 39202

RESULT 42
AC090822 181994 bp DNA HTG 27-MAY-2001
LOCUS Homo sapiens chromosome 11 clone CTD-2341A5 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 6 unordered pieces.
AC090822
AC090822
AC090822.2 GI:14210584
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FUILLTOP.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE 1 (bases 1 to 181994)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE Homo sapiens chromosome 11, clone CTD-2341A5

```

JOURNAL
REFERENCE
AUTHORS

Unpublished
2 (bases 1 to 181994)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Cammarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Fero, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J.,
Gardyna, S., Glende, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Laroque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
Mphahlele, R., Meldrum, J., Meneus, L., Mihova, T., Mleniga, Y.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkharn, P., Pierre, N., Pollara, V., Raymond, C., Retter, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Souarez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Trifilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 27, 2001 this sequence version replaced g1:13270691.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L12835

Center clone name: 2341.A.5

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179026 bases at least Q40
Consensus quality: 179649 bases at least Q30
Consensus quality: 179945 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 181494; sum-of-ctrls
Quality coverage: 10.9 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 609: contig of 609 bp in length
* 610 709: gap of 100 bp
* 710 1362: contig of 653 bp in length
* 1363 1462: gap of 100 bp
* 1463 8654: contig of 7192 bp in length
* 8655 8754: gap of 100 bp
* 8755 24527: contig of 15773 bp in length
* 24528 24627: gap of 100 bp
* 24628 82493: contig of 57866 bp in length
* 82494 82593: gap of 100 bp
* 82594 181994: contig of 99401 bp in length.

FEATURES
SOURCE

Location/Qualifiers
1. 181994
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="11"
/map="11"
/clone="CPD-2341A5"
/clone_11b="CITD Human BAC"
1. 609
/note="assembly_fragment"
710. 1362
/note="assembly_fragment"
1463. 8654
/note="assembly_fragment"
8755. 24527
/note="assembly_fragment"
24628. 82493
/note="assembly_fragment"
82594. 181994
/note="assembly_fragment"
BASE COUNT 56084 a 33254 c 33265 g 58889 t 502 others
ORIGIN

Query Match 1.5%; Score 20; DB 2; Length 181994;
Best Local Similarity 100.0%; Pired. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 12 caaatgtattttaacata 31
Db 70914 CAAATGTTATTTTAAACATA 70933

RESULT 43
AC026544/C
LOCUS
DEFINITION Homo sapiens chromosome 6 clone RP11-163122 map 6, WORKING DRAFT
ACCESSION AC026544
VERSION AC026544.2 GI:7528136
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 182807)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 6, clone RP11-163122
Unpublished
2 (bases 1 to 182807)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burnett, G.,
Campopiano, A., Cooke, P., Dearrellano, K., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domono, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D.,
Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczky, J.,
Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurt, A., McKernan, K., Mphahlele, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mleniga, Y., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Trifilio, J., Travers, M., Trifilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced g1:7283252.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 17232
Center clone name: 163.I.22
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165331 bases at least Q40
Consensus quality: 173794 bases at least Q30
Consensus quality: 177155 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 3.6 in Q20 bases; 3.8 in Q20 bases; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1088: contig of 1088 bp in length
*
* 1089 1188: gap of 100 bp
*
* 1189 2428: contig of 1240 bp in length
*
* 2429 2528: gap of 100 bp
*
* 2529 3878: contig of 1350 bp in length
*
* 3879 3978: gap of 100 bp
*
* 3979 5268: contig of 1290 bp in length
*
* 5269 5368: gap of 100 bp
*
* 5369 7753: contig of 2385 bp in length
*
* 7754 7853: gap of 100 bp
*
* 7854 9680: contig of 1837 bp in length
*
* 9691 9790: gap of 100 bp
*
* 9791 11570: contig of 1780 bp in length
*
* 11571 11670: gap of 100 bp
*
* 11671 13719: contig of 2049 bp in length
*
* 13720 13819: gap of 100 bp
*
* 13820 16027: contig of 2208 bp in length
*
* 16028 16127: gap of 100 bp
*
* 16128 18113: contig of 1986 bp in length
*
* 18114 18213: gap of 100 bp
*
* 18214 19867: contig of 1654 bp in length
*
* 19868 19967: gap of 100 bp
*
* 19968 21902: contig of 1935 bp in length
*
* 21903 22002: gap of 100 bp
*
* 22003 24557: contig of 2555 bp in length
*
* 24558 24657: gap of 100 bp
*
* 24658 27245: contig of 2588 bp in length
*
* 27246 27345: gap of 100 bp
*
* 27346 30656: contig of 3311 bp in length
*
* 30657 30756: gap of 100 bp
*
* 30757 33365: contig of 2609 bp in length
*
* 33366 33465: gap of 100 bp
*
* 33466 35816: contig of 2351 bp in length
*
* 35817 35916: gap of 100 bp
*
* 35917 38408: contig of 2492 bp in length
*
* 38409 38508: gap of 100 bp
*
* 38509 42615: contig of 4107 bp in length
*
* 42616 42715: gap of 100 bp
*
* 42716 47091: contig of 4375 bp in length
*
* 47092 47191: gap of 100 bp
*
* 47192 53079: contig of 5888 bp in length
*
* 53080 53179: gap of 100 bp
*
* 53180 56645: contig of 3466 bp in length
*
* 56646 56745: gap of 100 bp

```

FEATURES

SOURCE

```

* 56746 63020: contig of 6275 bp in length
* 63021 63120: gap of 100 bp
* 63121 71045: contig of 7925 bp in length
* 71046 71145: gap of 100 bp
* 71146 76850: contig of 5705 bp in length
* 76851 76950: gap of 100 bp
* 76951 85447: contig of 8497 bp in length
* 85448 85547: gap of 100 bp
* 85548 91907: contig of 6360 bp in length
* 91908 92007: gap of 100 bp
* 92008 98725: contig of 6718 bp in length
* 98726 98825: gap of 100 bp
* 98826 107166: contig of 8341 bp in length
* 107167 107266: gap of 100 bp
* 107267 115271: contig of 8005 bp in length
* 115272 115371: gap of 100 bp
* 115372 123500: contig of 8129 bp in length
* 123501 123600: gap of 100 bp
* 123601 133338: contig of 9738 bp in length
* 133339 133438: gap of 100 bp
* 133439 142253: contig of 8815 bp in length
* 142254 142353: gap of 100 bp
* 142354 155872: contig of 13519 bp in length
* 155873 155972: gap of 100 bp
* 155973 167753: contig of 11761 bp in length
* 167754 167853: gap of 100 bp
* 167854 182807: contig of 14954 bp in length.
*
Location/Qualifiers
1. 182807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-163122"
/clone_1fb="RP11 Human Male BAC"
1. 1088
/feature="assembly-fragment"
1189..2428
/feature="assembly-fragment"
2529..3878
/feature="assembly-fragment"
3979..5268
/feature="assembly-fragment"
clone_end:17
vector_side:right
5369..7753
/feature="assembly-fragment"
7854..9690
/feature="assembly-fragment"
9791..11570
/feature="assembly-fragment"
11671..13719
/feature="assembly-fragment"
13820..16027
/feature="assembly-fragment"
16128..18113
/feature="assembly-fragment"
18214..19867
/feature="assembly-fragment"
19968..21902
/feature="assembly-fragment"
22003..24557
/feature="assembly-fragment"
24658..27245
/feature="assembly-fragment"
27346..30656
/feature="assembly-fragment"
30757..33365
/feature="assembly-fragment"
33466..35816
/feature="assembly-fragment"
35917..38408
/feature="assembly-fragment"

```

```

misc_feature      38509..42615
                  /note="assembly-fragment"
misc_feature      42716..47091
                  /note="assembly-fragment"
misc_feature      47192..53079
                  /note="assembly-fragment"
misc_feature      53180..56645
                  /note="assembly-fragment"
misc_feature      56746..63020
                  /note="assembly-fragment"
misc_feature      63121..71045
                  /note="assembly-fragment"

```

```

Query Match      1.5%; Score 20; DB 2; Length 182807;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 119 aaaaagacattacttggtt 138
      |||
DB 153217 AAAAGACATTACTTGTT 153198

```

```

RESULT 44
AC080188      AC080188      183147 bp      DNA      HTG      26-OCT-2000
LOCUS          Homo sapiens chromosome 4 clone RP11-635F11, WORKING DRAFT SEQUENCE,
DEFINITION    40 unordered pieces.
ACCESSION     AC080188
VERSION       AC080188.4 GI:11024950
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 183147)
AUTHORS       Waterston, R.H.
TITLE         The sequence of Homo sapiens clone
REFERENCE     2 (bases 1 to 183147)
AUTHORS       Waterston, R.H.
TITLE         Direct Submission
JOURNAL       Submitted (28-SEP-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT       On Oct 26, 2000 this sequence version replaced gi:10945799.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0635L01
----- Summary Statistics -----
Sequencing vector: M13, 100%
Chemistry: Dye-primer ET, 100% of reads
Chemistry: Dye-terminator BigDye, 0% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 158849 bases at least Q40
Consensus quality: 167365 bases at least Q30
Consensus quality: 171079 bases at least Q20
Insert size: 185000; agarose-ftp
Insert size: 178294; sum-of-contigs
Quality coverage: 3.58 in Q20 bases; agarose-ftp
Quality coverage: 3.81 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
1 1171: contig of 1171 bp in length
* 1172: gap of unknown length
* 1272: contig of 1367 bp in length
* 2638: gap of unknown length
* 2738: gap of unknown length
* 3913: contig of 1175 bp in length
* 4013: gap of unknown length
* 6045: contig of 2032 bp in length
* 6046: gap of unknown length
* 6146: contig of 1767 bp in length
* 7912: gap of unknown length
* 8013: contig of 1943 bp in length
* 8013: gap of unknown length
* 9955: contig of 1943 bp in length
* 9956: gap of unknown length
* 10055: gap of unknown length
* 12248: contig of 2193 bp in length
* 12348: gap of unknown length
* 12349: contig of 2827 bp in length
* 15176: gap of unknown length
* 15275: contig of 2208 bp in length
* 17484: gap of unknown length
* 17583: contig of 2175 bp in length
* 19758: gap of unknown length
* 19759: gap of unknown length
* 19859: gap of unknown length
* 21922: contig of 2064 bp in length
* 21923: gap of unknown length
* 22022: gap of unknown length
* 22023: contig of 1857 bp in length
* 23879: gap of unknown length
* 23880: gap of unknown length
* 23979: contig of 2458 bp in length
* 25437: gap of unknown length
* 26438: contig of 2289 bp in length
* 26538: gap of unknown length
* 28826: contig of 2289 bp in length
* 28827: gap of unknown length
* 28927: gap of unknown length
* 31019: contig of 2093 bp in length
* 31020: gap of unknown length
* 33924: contig of 2805 bp in length
* 33925: gap of unknown length
* 34024: gap of unknown length
* 34025: contig of 3469 bp in length
* 37494: gap of unknown length
* 37594: gap of unknown length
* 40607: contig of 3014 bp in length
* 40608: gap of unknown length
* 40707: gap of unknown length
* 40708: contig of 3056 bp in length
* 43764: gap of unknown length
* 43764: gap of unknown length
* 43864: gap of unknown length
* 48779: contig of 4916 bp in length
* 48780: gap of unknown length
* 48880: gap of unknown length
* 53411: contig of 4531 bp in length
* 53511: gap of unknown length
* 57510: gap of unknown length
* 57510: contig of 4100 bp in length
* 57710: gap of unknown length
* 57711: gap of unknown length
* 58653: contig of 953 bp in length
* 58664: gap of unknown length
* 58763: gap of unknown length
* 58764: contig of 5135 bp in length
* 63899: gap of unknown length
* 63998: gap of unknown length
* 70378: contig of 6380 bp in length
* 70378: gap of unknown length
* 70478: gap of unknown length
* 76352: contig of 5774 bp in length
* 76353: gap of unknown length
* 76353: gap of unknown length
* 82864: contig of 6512 bp in length
* 82865: gap of unknown length
* 82964: gap of unknown length
* 82965: contig of 6633 bp in length
* 89597: gap of unknown length
* 89597: gap of unknown length
* 95761: contig of 6064 bp in length
* 95762: gap of unknown length
* 95861: gap of unknown length
* 95862: gap of unknown length
* 101602: contig of 5741 bp in length
* 101603: gap of unknown length
* 101603: gap of unknown length
* 108422: contig of 6720 bp in length
* 108423: gap of unknown length
* 108523: gap of unknown length
* 113599: contig of 5077 bp in length
* 113600: gap of unknown length
* 113600: gap of unknown length
* 121790: gap of unknown length
* 121794: contig of 8095 bp in length
* 121894: gap of unknown length
* 121894: gap of unknown length
* 130939: contig of 9045 bp in length
* 131039: gap of unknown length
* 138924: contig of 7885 bp in length
* 139024: gap of unknown length
* 139025: gap of unknown length
* 147837: contig of 8813 bp in length
* 147838: gap of unknown length
* 147937: gap of unknown length

```

```

* 147938 158381: contig of 10444 bp in length
* 158382 158481: gap of unknown length
* 158482 167102: contig of 8621 bp in length
* 167103 167202: gap of unknown length
* 167203 181613: contig of 14411 bp in length
* 181614 181713: gap of unknown length
* 181714 183147: contig of 1434 bp in length.
      location/qualifiers
1. 183147 "Homo sapiens"
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="4"
   /clone="RP11-635H1"
1. 1171
misc_feature /note="assembly_name:Contig13"
1272. 2638
misc_feature /note="assembly_name:Contig14"
2739. 3913
misc_feature /note="assembly_name:Contig16"
4014. 6045
misc_feature /note="assembly_name:Contig17"
6146. 7912
misc_feature /note="assembly_name:Contig18"
8013. 9955
misc_feature /note="assembly_name:Contig19"
10056. 12248
misc_feature /note="assembly_name:Contig20"
12349. 15175
misc_feature /note="assembly_name:Contig21"
15276. 17483
misc_feature /note="assembly_name:Contig22"
17584. 19758
misc_feature /note="assembly_name:Contig23"
19859. 21922
misc_feature /note="assembly_name:Contig24"
22023. 23879
misc_feature /note="assembly_name:Contig25"
23960. 26437
misc_feature /note="assembly_name:Contig26"
26538. 28826
misc_feature /note="assembly_name:Contig27"
28927. 31019
misc_feature /note="assembly_name:Contig28"
31120. 33924
misc_feature /note="assembly_name:Contig29"
34025. 37493
misc_feature /note="assembly_name:Contig30"
37594. 40607
misc_feature /note="assembly_name:Contig31"
40708. 43763
misc_feature /note="assembly_name:Contig32"
43864. 48779
misc_feature /note="assembly_name:Contig33"
48880. 53410
misc_feature /note="assembly_name:Contig34"
53511. 57610
misc_feature /note="assembly_name:Contig35
clone_end:sp6
vector_side:right"
57711. 58663
misc_feature /note="assembly_name:Contig12"
58764. 63898
misc_feature /note="assembly_name:Contig36"
63999. 70378
misc_feature /note="assembly_name:Contig37"
70479. 76252
misc_feature /note="assembly_name:Contig38"
76353. 82864
misc_feature /note="assembly_name:Contig39"
82965. 89597
misc_feature /note="assembly_name:Contig40"
89698. 95761
misc_feature /note="assembly_name:Contig41"

```

```

misc_feature 95862..101602
              /note="assembly_name:Contig42"
misc_feature 101703..108422
              /note="assembly_name:Contig43"
misc_feature 108523..113599
              /note="assembly_name:Contig44"

Query Match      1.5%; Score 20; DB 2; Length 183147;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 aaaaagacattacttggt 138
Db 100133 AAAAGACATTACTTGGT 100152

RESULT 45
AC079856/c 185421 bp DNA HTG 02-FEB-2001
LOCUS Homo sapiens chromosome RP11-521B4, WORKING DRAFT
DEFINITION SEQUENCE, 46 unordered pieces.
ACCESSION AC079856
VERSION AC079856.4 GI:12656844
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185421)
          Waterston,R.H.
          TITLE The sequence of Homo sapiens clone
          JOURNAL Unpublished
          AUTHORS 2 (bases 1 to 185421)
          Waterston,R.H.
          DIRECT SUBMISSION
          SUBMITTED (14-SEP-2000) Genome Sequencing Center, Washington
          UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          ON Feb 2, 2001 this sequence version replaced gi:11560300.

COMMENT ----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NE0521B04
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162590 bases at least Q40
Consensus quality: 169535 bases at least Q30
Consensus quality: 172901 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 180921; sum-of-ctnigs
Quality coverage: 3.10 in Q20 bases; sum-of-ctnigs
Quality coverage: 3.51 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1522: contig of 1522 bp in length
* 1523 1622: gap of unknown length
* 1623 2684: contig of 1062 bp in length
* 2685 2784: gap of unknown length
* 2785 4631: contig of 1847 bp in length
* 4632 4731: gap of unknown length

```

```

* 4732 6051: contig of 1320 bp in length
* 6052 6151: gap of unknown length
* 6152 7586: contig of 1435 bp in length
* 7587 7687 10059: gap of unknown length
* 7687 10059: contig of 2373 bp in length
* 10060 10159: gap of unknown length
* 10160 12215: contig of 2056 bp in length
* 12216 12315: gap of unknown length
* 12316 14002: contig of 1687 bp in length
* 14003 14102: gap of unknown length
* 14103 15350: contig of 1248 bp in length
* 15351 15450: gap of unknown length
* 15451 16663: contig of 1213 bp in length
* 16664 16763: gap of unknown length
* 16764 19156: contig of 2393 bp in length
* 19157 19256: gap of unknown length
* 19257 21199: contig of 1943 bp in length
* 21200 21299: gap of unknown length
* 21300 23500: contig of 2201 bp in length
* 23501 23600: gap of unknown length
* 23601 25776: contig of 2176 bp in length
* 25777 25876: gap of unknown length
* 25877 27788: contig of 1912 bp in length
* 27789 27888: gap of unknown length
* 27889 30449: contig of 2561 bp in length
* 30450 30549: gap of unknown length
* 30550 33166: contig of 2617 bp in length
* 33167 33266: gap of unknown length
* 33267 36506: contig of 3240 bp in length
* 36507 36606: gap of unknown length
* 36607 39573: contig of 2967 bp in length
* 39574 39674 42099: gap of unknown length
* 39674 42099: contig of 2426 bp in length
* 42100 42199: gap of unknown length
* 42200 45279: contig of 3080 bp in length
* 45280 45379: gap of unknown length
* 45380 48433: contig of 3054 bp in length
* 48434 48533: gap of unknown length
* 48534 51920: contig of 3387 bp in length
* 51921 52020: gap of unknown length
* 52021 55180: contig of 3160 bp in length
* 55181 55280: gap of unknown length
* 55281 57733: contig of 2453 bp in length
* 57734 57833: gap of unknown length
* 57834 60953: contig of 3100 bp in length
* 60954 61053: gap of unknown length
* 61054 64482: contig of 3429 bp in length
* 64483 64582: gap of unknown length
* 64583 67979: contig of 3397 bp in length
* 67979 68079: gap of unknown length
* 68080 71233: contig of 3154 bp in length
* 71234 71333: gap of unknown length
* 71334 75409: contig of 4076 bp in length
* 75410 75509: gap of unknown length
* 75510 80821: contig of 5312 bp in length
* 80822 80921: gap of unknown length
* 80922 85178: contig of 4257 bp in length
* 85179 85278: gap of unknown length
* 85279 89505: contig of 4221 bp in length
* 89506 89605: gap of unknown length
* 89606 94860: contig of 5255 bp in length
* 94861 94960: gap of unknown length
* 94961 100923: contig of 5963 bp in length
* 100924 101023: gap of unknown length
* 101024 107351: contig of 6328 bp in length
* 107352 107451: gap of unknown length
* 107452 113862: contig of 6411 bp in length
* 113863 113962: gap of unknown length
* 113963 118976: contig of 5014 bp in length
* 118977 119076: gap of unknown length
* 119077 124602: contig of 5526 bp in length
* 124603 124702: gap of unknown length
* 124703 131096: contig of 6394 bp in length

```

```

* 131097 131196: gap of unknown length
* 131197 138541: contig of 7345 bp in length
* 138542 138641: gap of unknown length
* 138642 146000: contig of 7359 bp in length
* 146001 146100: gap of unknown length
* 146101 154773: contig of 8673 bp in length
* 154774 154873: gap of unknown length
* 154874 163845: contig of 8972 bp in length
* 163846 163945: gap of unknown length
* 163946 175150: contig of 11205 bp in length
* 175151 175250: gap of unknown length
* 175251 185421: contig of 10171 bp in length.

FEATURES
source
1..185421"Homo sapiens"
/db_xref="taxon:9606"
/chromosome="PC1-11"
/clone="RP11-521B4"
1..1522
1..1522
/note="assembly_name:Contig13"
1623..2684
/note="assembly_name:Contig14"
2785..4631
/note="assembly_name:Contig17"
4732..6051
/note="assembly_name:Contig18"
6152..7586
/note="assembly_name:Contig19"
7687..10059
/note="assembly_name:Contig20"
10160..12215
/note="assembly_name:Contig21"
12316..14002
/note="assembly_name:Contig22"
14103..15350
/note="assembly_name:Contig23"
15451..16663
/note="assembly_name:Contig24"
16764..19156
/note="assembly_name:Contig25"
19257..21199
/note="assembly_name:Contig26"
21300..23500
/note="assembly_name:Contig27"
23601..25776
/note="assembly_name:Contig28"
25877..27788
/note="assembly_name:Contig29"
27889..30449
/note="assembly_name:Contig30"
30550..33166
/note="assembly_name:Contig31"
33267..36506
/note="assembly_name:Contig32"
36607..39573
/note="assembly_name:Contig33"
39674..42099
/note="assembly_name:Contig34"
42200..45279
/note="assembly_name:Contig35
vector_side:right"
45380..48433
/note="assembly_name:Contig36"
48534..51920
/note="assembly_name:Contig37"
52021..55180
/note="assembly_name:Contig38"
55281..57733
/note="assembly_name:Contig39"
57834..60953

Query Match 1.5%; Score 20; DB 2; Length 185421;

```



```

repeat_region      /rpt_family="POLY_A"
                    23650..23734
repeat_region      /rpt_family="L2"
                    complement(24237..24790)
repeat_region      /rpt_family="L2"
                    complement(24866..26486)
repeat_region      /rpt_family="L1MC3"
                    complement(26500..26578)
repeat_region      /rpt_family="L1MC/D"
                    complement(26586..26883)
repeat_region      /rpt_family="Aluy"
                    complement(26897..27432)
repeat_region      /rpt_family="L1MC/D"
                    complement(27664..27718)
repeat_region      /rpt_family="POLY_A"
                    complement(27720..28022)
repeat_region      /rpt_family="Alusg"
                    28425..28752
repeat_region      /rpt_family="L1MC4"
                    29233..29265
repeat_region      /rpt_family="AT_rich"
                    29277..29575
repeat_region      /rpt_family="Alusp"
                    complement(29883..29952)
repeat_region      /rpt_family="(GA)n"
                    complement(30056..30084)
repeat_region      /rpt_family="AT_rich"
                    31149..31315
repeat_region      /rpt_family="MLTIP"
                    32624..32743
repeat_region      /rpt_family="L2"
                    33420..33551
repeat_region      /rpt_family="Alusp"
                    33554..33854
repeat_region      /rpt_family="Aluy"
                    33855..34025
repeat_region      /rpt_family="Alusp"
                    36020..36320
repeat_region      /rpt_family="Alusp"
                    complement(36475..36582)
repeat_region      /rpt_family="L2"
                    complement(38770..38831)
repeat_region      /rpt_family="(TAA)n"
                    complement(38927..39036)
repeat_region      /rpt_family="(TAAA)n"
                    complement(39236..39304)
repeat_region      /rpt_family="MADE1"
                    complement(39359..39662)
repeat_region      /rpt_family="Alusg"
                    complement(40001..40026)
repeat_region      /rpt_family="AT_rich"
                    40601..40665
repeat_region      /rpt_family="MLT2B"
                    complement(40866..41165)
repeat_region      /rpt_family="Alusx"
                    complement(43174..43242)
repeat_region      /rpt_family="AT_rich"
                    complement(43995..44165)
repeat_region      /rpt_family="AluCo"
                    complement(44167..44490)
repeat_region      /rpt_family="Alusx"
                    complement(44495..44623)
repeat_region      /rpt_family="AluCo"
                    45064..45116
repeat_region      /rpt_family="POLY_A"
                    45217..45375
repeat_region      /rpt_family="MER5A"
                    46401..46610
repeat_region      /rpt_family="MER4D"
                    complement(46611..46666)

```

Query Match 1.5% Score 20 DB 9: Length 236281;
 Best Local Similarity 100.0% Pred. No. 20;

Matches	20; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY 265	gaataagaagagggacat	284			
Db 86019	GAATTAAGAAGAAGGACAT	86038			
RESULT 48					
AC005529					
LOCUS	AC005529	318488 bp	DNA	PRI	01-JUN-2000
DEFINITION	Homo sapiens chromosome 22q12 clone bk256d12, complete sequence.				
ACCESSION	AC005529				
VERSION	AC005529.7	GI:4646249			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Homo sapiens Chromosome 22q12 BAC Clone bk256d12 In MDR Region				
REFERENCE	Unpublished				
REFERENCE	2 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (25-AUG-1998) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	3 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (09-DEC-1998) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	4 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (05-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	5 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (07-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	6 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (09-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	7 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (10-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	8 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (14-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	9 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				
REFERENCE	Submitted (14-JAN-1999) Department Of Chemistry And Biochemistry,				
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,				
REFERENCE	OK 73019, USA				
REFERENCE	10 (bases 1 to 318488)				
REFERENCE	Ying, F., Pan, H. and Roe, B.A.				
REFERENCE	Direct Submission				

JOURNAL Submitted (16-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 11 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (19-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 12 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (24-JAN-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 13 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (18-FEB-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 14 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (27-FEB-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 15 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (04-MAR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 16 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (22-APR-1999) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 17 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (24-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 18 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (26-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 19 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (27-MAY-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 20 (bases 1 to 318488)

AUTHORS ying,f., Pan,h. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2000) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT On Apr 22, 1999 this sequence version replaced g1:4417314. Because these overlapping clones came from different libraries there are numerous instances of insertions, deletions, and single nucleotide polymorphisms in the overlapping regions below.

FEATURES

source 1. 318488

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="22q12"

/clone="BK256d12"

BASE COUNT 82006 a 78473 c 79291 g 78718 t

ORIGIN

Query Match 1.5% Score 20: DB 9: Length 318488;

Best Local Similarity 100.0% Pred. No. 21:

Matches 20: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 711 tgaattggtcagagattga 730

Db 162190 TGAATGTCAGAGATTGA 162209

RESULT 49

G55641/c 496 bp DNA STS 30-MAR-2000

LOCUS SHGC-100838 Human Homo sapiens STS genomic, sequence tagged site.

DEFINITION G55641

ACCESSION G55641.1 GI:6120810

VERSION

KEYWORDS STS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 496)

AUTHORS Olivier,M. and Cox,D.R.

TITLE Unpublished, Olivier, M., Cox, D.R. (2000)

JOURNAL Unpublished (2000)

COMMENT Contact: Michael Olivier, David R. Cox

Stanford Human Genome Center

Stanford University School of Medicine

4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA

Tel: (650) 320-5800

Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu

Primer A: CTTGATGAGGAGACTTTCAGATG

Primer B: ACGCAACATCATCATACATCGTC

STS size: 282

PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds

Annealing: 60 degrees C for 30 seconds

Polymerization: 72 degrees C for 23 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9700

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Amplifing Gold Polymerase: 0.07 units/uL

Total Vol: 5 uL

Buffer: MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 8.3

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed

FEATURES and developed at the Stanford Human Genome Center.

SOURCE

Location/Qualifiers
1. 496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1"
/clone_lib="Human"

STS

primer_bind

180..461

180..202

complement(439..461)

BASE COUNT

181 a 72 c 112 g 131 t

ORIGIN

Query Match

1.4%; Score 19; DB 11; Length 496;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

549 ctccaatatttcattgct 567

Db 232 CTCGATATTTTCATGCT 214

RESULT 50

AF217223/C

AF217223 1087 bp DNA

VR 03-AUG-2000

LOCUS

Influenza B virus (B/Netherlands/580/89) hemagglutinin gene,

DEFINITION

partial cds.

ACCESSION

AF217223

VERSION

AF217223.1 GI:7650317

KEYWORDS

Influenza B virus (B/Netherlands/580/89).

SOURCE

Influenza B virus (B/Netherlands/580/89).

REFERENCE

Viruses; ssRNA negative-strand viruses; Orthomyxoviridae; Influenza

AUTHORS

1 (bases 1 to 1087)
Osterhaus, A.D., Rimmelzwaan, G.F., Martina, B.E., Bestebroer, T.M. and

TITLE

Fouchier, R.A.

JOURNAL

Influenza B virus in seals

MEDLINE

Science 288 (5468), 1051-1053 (2000)

PUBMED

10807575

REFERENCE

2 (bases 1 to 1087)

AUTHORS

Fouchier, R.A., Rimmelzwaan, G.F., Martina, B.E., Bestebroer, T.M. and

TITLE

Direct Submission

JOURNAL

Submitted (20-DEC-1999) Virology, Erasmus University, Dr.

FEATURES

Molewaterplein 50, Rotterdam 3015 GE, Netherlands

SOURCE

Location/Qualifiers

CDS

1. 1087
/organism="Influenza B virus (B/Netherlands/580/89)"
/strain="B/Netherlands/580/89"

FEATURES

<1..>1087
/note="HA1 domain"

CDS

/codon_start=2
/product="hemagglutinin"

FEATURES

/protein_id="AAF65990.1"

FEATURES

/db_xref="GI:7650318"

FEATURES

/translation="MKATIVLMVTSNADRICGTGTSNSPHVVKTAOGENVTVG

FEATURES

IPITTPYKSHRANKGKTRGKLCPCNCLCTDDVVALRPMKCVGTIPSAKASIIHAY

FEATURES

RPVTSQCFPIHMDRIKIRQIPNLRLGYNIRLSTQVINAERKAPGPIRLGTSGSCP

FEATURES

VTSRNGEFAIMAAVPRDKRTAINPLTVEVPYICNGEDQITVWGFHSNDKTKOMKLY

FEATURES

GDSNPKETSSANGVTTHVYSQIGFPPNOTEDGGLPQSGRIYVDYMWOKPGKGTIVY

FEATURES

ORGVLLPKQWMCAGRSKVKYLSLPLIGADCLHEKYGGLNKSFKPYTGEHAKAIGNC

FEATURES

PIWKTPIKLANGTKYRPPAKLIKRGF"

BASE COUNT

381 a 243 c 227 g 236 t

ORIGIN

QY 764 tgaaccccccaacagtaa 782
Db 621 TGAACCCCCCAACAGTAA 603

Search completed: May 2, 2002, 13:18:43
Job time: 9727 sec

Query Match 1.4%; Score 19; DB 14; Length 1087;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Fri May 3 10:57:48 2002

us-09-645-192-1.0110.rge

Page 52

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2002, 07:45:15 ; Search time 28.73 Seconds

(without alignments)
2306.348 Million cell updates/sec

Title: US-09-645-192-2

Perfect score: 2389
Sequence: 1 MKIFKCYFHTLQOKVFILF.....DWITLPSKELFMDRLTTTS 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2389	100.0	453	4	Q9P109	Q9P109 homo sapien
2	868	36.3	438	4	O95395	O95395 homo sapien
3	866.5	36.3	428	11	O64165	O64165 rattus norv
4	848	35.5	440	12	O912K2	O912K2 bovine herp
5	847	35.5	398	11	Q9D8A3	Q9D8A3 mus musculu
6	844.5	35.3	428	11	O35981	O35981 mus musculu
7	842	35.2	440	12	O99CM3	O99CM3 bovine herp
8	645.5	27.0	356	11	O9D2A8	O9D2A8 mus musculu
9	508.5	21.3	314	4	O9HCY8	O9HCY8 homo sapien
10	326	13.6	445	5	O19730	O19730 caenorhabdi
11	323	13.5	425	5	O02234	O02234 caenorhabdi
12	315.5	13.2	454	5	Q19729	Q19729 caenorhabdi
13	296	12.4	401	5	O02314	O02314 caenorhabdi
14	294	12.3	472	5	O20406	O20406 caenorhabdi
15	275	11.5	402	5	O02315	O02315 caenorhabdi
16	245	10.3	459	5	O02309	O02309 caenorhabdi
17	231	9.7	876	5	O9W034	O9W034 dirosophila
18	229.5	9.6	454	5	O45866	O45866 caenorhabdi
19	224.5	9.4	864	11	Q9EP10	Q9EP10 rattus norv

20	222.5	9.3	322	2	Q9A4H4	Q9A4H4 caulobacter
21	220.5	9.2	489	5	O21796	O21796 caenorhabdi
22	219.5	9.2	470	5	O16776	O16776 caenorhabdi
23	219.5	9.2	486	5	O93643	O93643 caenorhabdi
24	219.5	9.2	865	4	O9H1B5	O9H1B5 homo sapien
25	214.5	9.0	616	11	Q9EP10	Q9EP10 mus musculu
26	213.5	8.9	821	11	Q9EP11	Q9EP11 rattus norv
27	213.5	8.9	789	4	O9H1B6	O9H1B6 homo sapien
28	211.5	8.9	827	11	Q9EP11	Q9EP11 mus musculu
29	203.5	8.5	472	5	O76634	O76634 caenorhabdi
30	202.5	8.5	753	5	O02312	O02312 caenorhabdi
31	197.5	8.3	448	10	Q9Z0Z7	Q9Z0Z7 arabisdopsis
32	194.5	8.1	513	5	O9TVP0	O9TVP0 caenorhabdi
33	193.5	8.1	478	5	O9XXL2	O9XXL2 caenorhabdi
34	188	7.9	467	5	O9U3H9	O9U3H9 caenorhabdi
35	187	7.8	447	10	O9LR71	O9LR71 arabisdopsis
36	182.5	7.6	513	5	O9XTM4	O9XTM4 caenorhabdi
37	181	7.5	424	10	O9LFE0	O9LFE0 arabisdopsis
38	178	7.5	367	5	O22481	O22481 caenorhabdi
39	175.5	7.3	406	10	O9LNM5	O9LNM5 arabisdopsis
40	172.5	7.2	378	10	O9S6S9	O9S6S9 arabisdopsis
41	169.5	7.1	378	10	O9ASZ7	O9ASZ7 arabisdopsis
42	168	7.0	395	10	O9C9A1	O9C9A1 arabisdopsis
43	166.5	7.0	434	10	O9LFD0	O9LFD0 arabisdopsis
44	151	6.3	447	10	O9FLD7	O9FLD7 arabisdopsis
45	139	5.8	402	10	O9AR03	O9AR03 oryza sativ

ALIGNMENTS

RESULT 1
ID Q9P109 PRELIMINARY: PRT: 453 AA.
AC Q9P109;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CORE 2 BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE 3.
GN C2GNT3.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20219156; PubMed-10753916;
RA Schwienbeck T., Yen J.C., Levery S.B., Keck B., Merix G.,
RT van Kessel A.G., Fukuda M., Clausen H.;
RT "Control of O-glycan branch formation. Molecular cloning and
RT characterization of a novel thymus-associated core 2 beta1,6-N-
RT acetylglucosaminyltransferase.";
RT J. Biol. Chem. 275:11106-11113(2000).
DR EMBL: AF132035; AAF63156.1;
DR InterPro: IPR003406; Branch.1.
DR Pfam: PF02485; Branch.1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 453 AA; 53052 MW; B43794D4427F41CA CRC64;

Query Match 100.0%; Score 2389; DB 4; Length 453;
Best Local Similarity 100.0%; Pred. No. 1e-173;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKIFKCYFHTLQOKVFILFMTLSLKLNVRLFPQKDIYVERSLSTSPVRRNY 60
DB 1 MKIFKCYFHTLQOKVFILFMTLSLKLNVRLFPQKDIYVERSLSTSPVRRNY 60
QY 61 THVKDVRVENVSGIYVEPELIGKSLERRRDIIDDDVAVMTSCDIYQTLRGTA 120
DB 61 THVKDVRVENVSGIYVEPELIGKSLERRRDIIDDDVAVMTSCDIYQTLRGTA 120
QY 121 OKLVSEKESFPIAVSLVYHKDAIWERLHAIYNOHNIYCIHYDRKAPDFKVAAMNLA 180

```

Db 121 OQLVSKSEKSPFLAYSLVHKDAIMVERLHAITNOHTICIHDKRAPDPTFYAMNNLA 180
QY 181 KCFSNIFIAKLEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELV 240
Db 181 KCFSNIFIAKLEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELV 240
QY 241 SELKINGANMLETVKPNKSLERFTYHHLRVPYEVYKLPRTINSKAPPHNIOIFV 300
Db 241 SELKINGANMLETVKPNKSLERFTYHHLRVPYEVYKLPRTINSKAPPHNIOIFV 300
QY 301 GSAFVLISQAFVYKVIENNSIVQDFEFAWSKDTYSPDEHFMTLLIRVPGIPGEISRAQDVS 360
Db 301 GSAFVLISQAFVYKVIENNSIVQDFEFAWSKDTYSPDEHFMTLLIRVPGIPGEISRAQDVS 360
QY 361 DLQSTRLVKNWYEGFFPSCGSHLRVCIYGAELRMLIKDGHMFANKFDSKVPIL 420
Db 361 DLQSTRLVKNWYEGFFPSCGSHLRVCIYGAELRMLIKDGHMFANKFDSKVPIL 420
QY 421 IKCLAEKLEEQORDWITLPSSEKLFMDRLTTTS 453
Db 421 IKCLAEKLEEQORDWITLPSSEKLFMDRLTTTS 453

```

RESULT 2
095395 PRELIMINARY; PRT; 438 AA.

```

AC 095395;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE BETA-1,6-N-ACETYLGLUCOSAMINYLTRANSFERASE.
GN C2/4GNF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
CX [1]
RN TISSUE-BRAIN;
RC MEDLINE=9915671; PubMed=9915662;
RA Yeh J.C., Ong E., Fukuda M.;
RT "Molecular cloning and expression of a novel beta-1, 6-N-
RT acetylglucosaminyltransferase that forms core 2, core 4, and I
RT branches.";
RL J. Biol. Chem. 274:3215-3221(1999).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99143102; PubMed=9986882;
RA Schiavetek T., Nomoto M., Levery S.B., Merx G., van Kessel A.G.,
RT "Control of O-glycan branch formation. Molecular cloning of human cDNA
RT encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core
RT 2 and core 4.";
RL J. Biol. Chem. 274:4504-4512(1999).
DR EMBL; AF102542; AAD10824.1; -
DR EMBL; AF038650; AAD21525.1; -
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
DR Transferrase; Glycosyltransferase.
SQ SEQUENCE 438 AA; 50863 MW; 1FF0A1B451C98407 CRC64;

```

Query Match 36.3%; Score 868; DB 4; Length 438;
Best Local Similarity 43.0%; Pred. No. 4,6e-58;
Matches 182; Conservative 71; Mismatches 146; Indels 24; Gaps 9;

```

QY 23 LMLSLKLKLVN-----KRLPFQKITYLVESLSISPVVRN-RTHVADERYVYVNSG 75
Db 13 LMLLGYMLLATVALKLSFRLKCDSDHGLRESQSQYCRALINFLKLPKRSINCSG 72
QY 76 IV--EQQPL--EIGKSLERIRRDIDLEDVVAMTSDCDIYQTLRGYAKIVSEKSF 131
Db 76 IV--EQQPL--EIGKSLERIRRDIDLEDVVAMTSDCDIYQTLRGYAKIVSEKSF 131

```

```

Db 73 VTRGDQEAIVLQALINNLVEYKKR--EPFTDHYSLTRDCEHKAERKEIOFPLSKEEYEF 131
QY 132 PIAYSLVYVKDAIMVERLHAITNOHTICIHDKRAPDPTFYAMNNLA 180
Db 132 PIAYSLVYVKDAIMVERLHAITNOHTICIHDKRAPDPTFYAMNNLA 180
QY 132 PIAYSVYIHEKTEENFERLRAVAYAPONICVYDEKSPETFEAKAISCPNPFVIAK 191
Db 132 PIAYSVYIHEKTEENFERLRAVAYAPONICVYDEKSPETFEAKAISCPNPFVIAK 191
QY 192 LEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
Db 192 LEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
QY 192 LVAVVYASMSRVQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
Db 192 LVAVVYASMSRVQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELKINGANM 251
QY 252 LETVKNPNSKLERFTYHHLRVPYEVYKLPRTINSKAPPHNIOIFV 300
Db 252 LETVKNPNSKLERFTYHHLRVPYEVYKLPRTINSKAPPHNIOIFV 300
QY 312 VKYTFNNNSIVQDFEFAWSKDTYSPDEHFMTLLIRVPGIPGEISRAQDVS 360
Db 312 VKYTFNNNSIVQDFEFAWSKDTYSPDEHFMTLLIRVPGIPGEISRAQDVS 360
QY 366 WOCHGEDIDKGAEPYAPCSGIIHORALICVYAGDLMNMLQNHILLANKFDPKVDNALQCLE 425
Db 366 WOCHGEDIDKGAEPYAPCSGIIHORALICVYAGDLMNMLQNHILLANKFDPKVDNALQCLE 425
QY 426 EKL 428
Db 426 EKL 428

```

RESULT 3
064165 PRELIMINARY; PRT; 428 AA.

```

AC 064165;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ENZYMAIC GLYCOSYLATION-REGULATING GENE PROTEIN.
DE Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
CX [1]
RN TISSUE-BRAIN;
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HEART;
RX MEDLINE=96013636; PubMed=7560067;
RA Nishio Y., Warren C.E., Buczek-Thomas J.A., Rulis J., Koya D.,
RT "Identification and characterization of a gene regulating enzymatic
RT glycosylation which is induced by diabetes and hyperglycemia
RT specifically in rat cardiac tissue.";
RL J. Clin. Invest. 96:1759-1767(1995).
DR EMBL; S79797; AAB35697.2; -
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
SQ SEQUENCE 428 AA; 49826 MW; AE25A18172897AA8 CRC64;

```

Query Match 36.3%; Score 866.5; DB 11; Length 428;
Best Local Similarity 42.3%; Pred. No. 5,8e-58;
Matches 183; Conservative 73; Mismatches 136; Indels 41; Gaps 11;

```

QY 28 LKLKLVNRLP--PQNDIVLV-EXSLSTSPFVRNR-----YTHVK---DEVRYEYVNS 74
Db 1 MLRNLFRRLRFLSYPTKTYFNVVLSTLTFGVRHOKPEFVSVSHLESDDDNSNVNCT 60
QY 75 GIVQDEPLEIGK-SLEI-----RRDIIIDEDDVVAMTSDCDIYQTLRGYAKIV 124
Db 61 KVLQGDDELEIQKLEILITVQFKRPRR-----TPHDYINMRDCASFRTIKTYMEPL 114
QY 125 SKEKSPFIAYSLVYVKDAIMVERLHAITNOHTICIHDKRAPDPTFYAMNNLA 184
Db 115 TKEKSPFIAYSLVYVKDAIMVERLHAITNOHTICIHDKRAPDPTFYAMNNLA 184
QY 185 NIFIAKLEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELK 244
Db 185 NIFIAKLEAVEYAHISRLQADINCLSDLLKSSIQMKYVINLCGDFPLKSNELVSELK 244

```

```

Db 175 NFEVASQLESVYASWSRYKADLNCMDLYRMANKVYLINCGMDFPIKTNLEIYRKAK 234
OY 245 KUNGAMLETVKPPNSKLEFRTHHEHLRMPVEYVYLIRTNISKAPRHNQIIFVGSAY 304
Db 235 SFTGENSELSETERKPPKKEERAK-----KRYTYDGKL-INTGVYKQPLKPLIFSGSAY 288
OY 305 FVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGIPGEISRSAO-DVSDIQ 363
Db 289 FVYTRREYGVYLENNKNIKFMEMWAQDTYSPDELMATIORIPFPGSLPSHRYDLSDMN 348
OY 364 SKRTIKMNYEGEF-----YPSCTGSHLRVCYGAELRMLIKDHWGANKFDSKVPD 418
Db 349 AVARFVKMQYFEGDVSNGADYPPGSGVHVASCVGVGDLSSMLKRNHFFANKFMDVDP 408
OY 419 ILIKLAELKEEQ 431
Db 409 FALQCEHLRHK 421

RESULT 4
OY1K2 PRELIMINARY; PRT; 440 AA.
AC OY1K2;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE BETA-1,6-N-ACETYLGALUCOSAMINYLTRANSFERASE.
GN BOREF3-4.
OS Bovine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OC NCBI_TaxID=10385;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V. TEST;
RX MEDLINE=20283898; PubMed=10811884;
RA Vanderplasschen A., Markine-Gorlaynoff N., Lomonte P., Suzuki M.,
RA Hiraoka N., Yeh J.-C., Bureau F., Williams L., Thiry E., Fukuda M.,
RA Pastoret P.-P.;
RT "A multipotential beta-1,6-N-acetylglucosaminyl-transferase is encoded
RT by Bovine herpesvirus type 4."
RL Proc. Natl. Acad. Sci. U.S.A. 97:5756-5761(2000).
DR EMBL; AF231105; AAF72001.1;
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
KM Transferase; Glycosyltransferase.
SQ SEQUENCE 440 AA; 50701 MW; 71B9C3B6B4A949BD CRC64;

Query Match 35.5%; Score 848; DB 12; Length 440;
Best Local Similarity 44.0%; Pred. No. 1.5e-56;
Matches 168; Conservative 72; Mismatches 124; Indels 18; Gaps 6;
OY 60 YTHVDEVREYVNCSGY--EOEPL--EIGKSLERRADIDLEDDVVAAMTSDCIYOT 115
Db 60 YNSILPKRISNCIGIRGDEAYVQALDNLVEYKKR-PLPDTYLTNTTROCERKA 118
OY 116 LRGYAQKLVSEKESFPIAYSLVHKDAIWERLTHAIYNQHNICYIYDRKAPDFKVA 175
Db 119 QKRFIOPLSKRELDFPIAYSMVYHEKLENERLRAVYAQNIYCVHADVKSDETFKEA 178
OY 176 MNKLKCFNSIRIASKLEAVEYAHISRLQADINCLSDLLKSSIOKKYVINCGOFPPLKS 235
Db 179 VVAIISCFPNVMAKSLPVVYASMSRQADINCMEDLLQSSVYKYLINTCGIDFPKTI 238
OY 236 NFEVSELKINGANMLETVPKNSKLERFTYHHELRNRYEYVKLPRTNISKAPRPN 295
Db 239 NAEVYLAKMLKLGKXMSMESEVSESKKMKMYREYVTDLTLPYTSKI-----KDPDPDN 291
OY 296 IOIFGSAFVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGIPGEI-SR 354
Db 292 LRFITGNATFVASRAFAVQAVLDNPKRSQILVENYKDTYSPDEHLMTLQRAAPMPSVSH 351

```

```

OY 355 SAQDVSQSTRVLYKNYEGEF-----YPSCTGSHLRVCYGAELRMLIKDHWGANKF 409
Db 352 PKYHISDMTALARLYKMQVHEGDSMGAPVAPCSGIRHRAICIVAGADLYWILQNHLLA 411
OY 410 NKFDKVPILIKLAELKEEQ 431
Db 412 NKFDPRVDNVLQCLEEYLRHK 433

RESULT 5
OY9DA3 PRELIMINARY; PRT; 398 AA.
AC OY9DA3;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE 2010013H22RIK PROTEIN.
GN 2010013H22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamana I.,
RA Satoh T., Okazaki Y., Gojodori T., Sono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.O., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.B., Kamija M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashitaki Y.;
RT "Functional annotation of a full-length mouse CDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008234; BAB25548.1;
DR MGI; MGI:191327; 2010013H22RIK.
DR InterPro; IPR003406; Branch.
DR Pfam; PF02485; Branch; 1.
SQ SEQUENCE 398 AA; 46000 MW; 275ABE93909CC4B6 CRC64;

Query Match 35.5%; Score 847; DB 11; Length 398;
Best Local Similarity 42.4%; Pred. No. 1.6e-56;
Matches 162; Conservative 79; Mismatches 123; Indels 18; Gaps 5;
OY 60 YTHKDEVREYVNCSGYEOEPLCK-----SLFIRRODIDLEDDVVAAMTSDCIYOT 115
Db 18 YKTLKPAKISNCSGIRGEQKAVTQALNNLEIKKQQL-FTEDLYLMTADCHHFT 76
OY 116 LRGYAQKLVSEKESFPIAYSLVHKDAIWERLTHAIYNQHNICYIYDRKAPDFKVA 175
Db 179 VVAIISCFPNVMAKSLPVVYASMSRQADINCMEDLLQSSVYKYLINTCGIDFPKTI 238
OY 236 NFEVSELKINGANMLETVPKNSKLERFTYHHELRNRYEYVKLPRTNISKAPRPN 295
Db 137 VRAVLSQFPPNVFASKLAVSVYASMSRQADINCMEDLLQSSVYKYLINTCGIDFPKTI 196
OY 296 IOIFGSAFVLSQAFVKYIFNNSIVODFANSKDTYSPDEHFMTLRVPGIPGEI-SR 354
Db 197 NAEVYLAKMLKLGKXMSMESEVSESKKMKMYREYVTDLTLPYTSKI-----MTSKRRTPPNN 249

```

Fri May 3 10:57:56 2002

us-09-645-192-2.rspt

Page 4

[illegible]

RESULT	6
035981	
ID	035981
AC	035981
PRELIMINARY:	
PRT:	428 AA.
DT	01-JAN-1998 (TREMBLrel. 05, Created)
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	ACETYLGLUCOSAMINYL TRANSFERASE 1, CORE 2 (BETA-1,6-N-
DE	ACETYLGLUCOSAMINYLTRANSFERASE).
GN	GCRT1.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
NCBI_TaxId	10090;
[1]	
SEQUENCE FROM N.A.	
RC	STRAIN=BALE/C; TISSUE=SUBMAXILLARY GLAND, AND KIDNEY;
RC	MEDLINE=98001705; PubMed=9341170;
RA	Sekine M., Nara K., Suzuki A.;
RT	"Tissue-specific regulation of mouse core 2 beta-1,6-N-
RT	acetylglucosaminyltransferase.";
RL	J. Biol. Chem. 272:2246-2252(1997).
DR	EMBL: D87335; BAA2299.1; -
DR	EMBL: D87332; BAA2298.1; -
DR	MGI: MGI.95676; Genti.
DR	Interpro: IPR003406; Branch.
DR	Pfam: PF02485; Branch: 1.
KW	Transferase; Glycosyltransferase
SO	SEQUENCE 428 AA; 49831 MW; PF010A6B23B92C CRC64;

```

Query Match          35.33:  Score 94.45:  DB 11:  Length 428:
Best Local Similarity 42.44:  Pred. No. 2.7e-56:
Matches 183:  Conservative 71:  Mismatches 139,  Indels 39,  Gaps

QY      28  LKILINVERLF--POKDIY-LVEYSLSSTSPFNRRYTHVDE--VRY-----EV 71
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      1  MLRNLFRRRLFSQCPKRYKYMLLVSLITFSVL---IHQKPEFYSVHLELADDDPSNV 57
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      72  NCSGIEQDPELEIG-----SLERRDITDIEDDDVAVANTSCDIYQTLRGYAOKYS 125
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      58  NQTKLLODDPEELQVKVLELITVOFKRP--RMTPHYIIMNTBCASFIRKRYIEPLT 115
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      126  KEKSEFPTAYSLVYKADIAINVERLIHAHYONHNTICHYDRKADPEKVMANMLACFSN 185
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      116  KEVEGFPRLSYVHNKIKEMLDRLIRATYMQNTCYCHDRKAEESLAAVQIGASCFDN 175
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      166  IFIASKLEAVEYAHISRLQADNLCSIDLKSSIQMKYVINLCGDFELKSNFELVSELK 245
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      176  VEVASQLESVYVYASMSVPKADNLCKMDLRNNAMMKYILNLGMDPEIFIKTLEIVRKLK 235
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      246  LAGANMLEIVKRPNSKLERFTYHNHLRVRVPELYKLPRTNISKAEAPRHNIQIFVGAYF 305
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      236  SPAENNELTEKMPKPEERWK-----KRYAVVDGL-TNGIYVAPRLPTLPESGAYF 289
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      306  VLSQAFQIKFNSYIOFFAFMSKIDPDEHFAVATILRYGIGIEI--SRSAODPLOS 364
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      290  VYTRVEYGVLENEINIQKLMENAAQOTYSPDEFIVATIORIDVYGCSFPSSKUYDLSDMA 349
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      365  KTRLVKWNYYBGF-----YBSCIGSHLRVSVICGADELRLIKDGHMFANKEDSKVDPI 419

```

Db	350	IARFVKWQYIFEGGSDVNSGAPFPCSGVHYRVCYFGAGDLSMLRLQHHLTFANKEDIDVDPF	409
QY	420	LIKCIAERLEQ	431
Db	410	ATQCIDEHLRRK	421

RESULT	7		
099CM3			
ID	099CM3	PRELIMINARY:	PRN: 440 AA.
AC	099CM3		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	VIRAL BERA-1,6-N-ACETYLGLYCOSAMINYLTRANSFERASE.		
OS	Bovine herpesvirus 4.		
CC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
CC	Gammapherpesvirinae.		
OX	NCBI_TaxID=10385;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=2053805; PubMed=1152491;		
RA	Zimmermann W., Broli H., Ehlers B., Buhk H.-J., Rosenthal A.,		
RA	Goltz M.;		
RT	"Genome Sequence of Bovine Herpesvirus 4, a Bovine Rhadinovirus, and		
RT	Identification of an Origin of DNA Replication.",		
RL	J. Virol. 75:1186-1194(2001).		
DR	EMBL: AF318573; AK07999.1; -		
KM	Transferase; Glycosyltransferase.		
SO	Transferase; 440 AA; 50763 KM; C08B3EAB8FDB8DD9 CRCC4;		

	Query Match Similarity	35.2%	Score 842;	DB 12;	Length 440;
	Best Local Similarity	43.7%	Pred. No. 4,4e-56;		
Matches	167;	Conservative	72;	Mismatches	125; Indels 18; Gaps 6;
Qy	60 YTHVXDEVRYEVNCSGIV--EQEPL--ELIGSLERIRDDI IDEEDDVAAMTSDCDIYOT	115			
Dd	60 YMSLKPAKRISINGSGITGDDEAVQAALDNLEVKKKR-SPLTGTYLNIIRDCERFFA	118			
Qy	116 LKNGAQKYSKREKSPPRIYASLVVHNDALMVEBLIAINOHNIYOIHVDRAKPDEFYA	175			
Dd	119 QGRFIOFPFLSKELDPIYSVMWHEKIKINFEBLLAAVAPONTICVHDVASPEFKFA	178			
Qy	176 MNMLAKFSNFIFASKLEAVEYAHISRLOADNCLSIDLSKIOMKVYNLCGODFELKS	235			
Dd	179 VKATISCENVFMASTLEVUYVASMRVGADLMCMEDLDLOSSVPWKYLLINTCGTDPIKT	238			
Qy	236 NRELVSCLKKANMLEVVRPNKSLSEFTYHNHLRRVREYVKLRPITNISKEAPRN	295			
Dd	239 NMEMVALKLMLKGNSMEESVPSSESKNNMKWTRYETDTLPYTSKM-----KDRPDN	291			
Qy	296 IOIFVSAAYFLVSOAEVKYTFNNSIQODEFAASKIDVSPDENFWALLIRVGPIGEIR-SR	354			
Dd	292 LPREFGNAAFVNASRAYVGHLYDLDPKQRQLRVEMWKDIYSPDEHLMATLOARPMWPBSH	351			
Qy	355 SAODVSDLOSRTRLVKNWTYEGRF----YPSTGSGLHSVCITYGAABLRLILKDGHWA	409			
Dd	352 PKYHIDMDTAIALVLWMQHNEGDSMGARYAPCDSGIHRRAICTYGAGDWLIIONHHLLA	411			
Qy	410 NKPFDSKVPFLIKCLAEKLEEQ	431			
Dd	412 NKFDPRVDDNVLOCLEYLRLHK	433			
RESULT	8.				
ID	Q9D2A8				
AC	Q9D2A8:	PRELIMINARY:	PRT;	356 AA.	
DT	01-JUN-2001 (TEMBRel. 17, Created)				
DT	01-JUN-2001 (TEMBRel. 17, Last sequence update)				
DT	01-JUN-2001 (TEMBRel. 17, Last annotation update)				

Query Match 12.4%; Score 296; DB 5; Length 401

Query Match	12.3%;	Score 294;	DB 5;	length 472;
Best Local Similarity	21.8%;	Pred. No. 2.3e-14;		

Matches 103: Conservative 83; Mismatches 184; Indels 102; Gaps 17;

QY 32 LNWRR-----LEPQDIYL-----VEYSLSTSPFVRRNRYHVDENVYNGSGIYE 78
 Db 24 MNVRRFKLISOFIYSSFTYIVLRIGFVDYSL-----VPENRK-----LE 63
 QY 79 QEPLEIKSLLE-IRRDDI-----IDLEDDVYAMTSDCDIYQTLRGVAOKIVSKRE 127
 Db 64 NIPLECSVINGTENRKRKISARQMDWHFDWVEHEINRKNVCSTIDKYINFRIPSSRE 123
 QY 128 EKSPPIAYSLVNHKDAIMVERLHAIFYNQHNICYIHYDRKAPDTEKYAMNNLAKCESN-- 185
 Db 124 EAEPFLAVGLVYKTYIYOVLQOMSLFYQPOHMFICITVDOSPNKYSVIALPSCSPNMH 183
 QY 186 IFISKLEAVEXAHISRLQADNLCLSLKSSIQMKRYVINLGGDFPLKSNFELYSCLK 245
 Db 184 VFIG---EPSQMGSPGILKNYITCFNMLSKSKQKWKYQILSGIDLEIRNLKLVRFKA 240
 QY 246 LGANMLETVPKPSKLERFTYHHLRVRVYEVYKLPRTNISKAPDPHNIQIFVGSAYF 305
 Db 241 LNGS-----NMVDVSTF---EVDRY-----KMEGVLP--MPYKSSMSY 276
 QY 306 VLSCAFVYIYNNSIYODFFAMSKDTPSPDEHFMTLIRVPG---IPGEISRSADQVSD 362
 Db 277 VVPRGADYLLISPRVQKLKYLKTYLWLPDESFWSTYVGSFALLPVGST--RYRDLTLM 334
 QY 363 QSKTRL-----VKMNYEGF---FYPSCTGSHRSVCITYGAELFWLIKDGHW 408
 Db 335 RKNFLRPEYNTVNSISTSYIGRYQVGMQKECFGKVKYDSCYGFEDIEIMTREL 394
 QY 409 ANKPEKSDPILIKCLAEKLEEQ-----RDWTLTPEKLFMDRNT 450
 Db 395 AHKLYLEPQAFQMKLEKVRRLSLSPDALFSAISISQMTVELYQCKAT 446

RESULT 15
 002315 PRELIMINARY; PRT; 402 AA.

AC 002315.
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE T15D6.3 PROTEIN.
 GN T15D6.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 DR EMBL: Z83125; CAB05621.1; -
 DR InterPro: IPR003406; Branch.
 DR Pfam: PF02485; Branch; 1.
 SQ SEQUENCE 402 AA; 46135 MW; F6AD721976703307 CRC64;

Query Match 11.5%; Score 275; Db 5; Length 402;
 Best Local Similarity 24.0%; Pred. No. 5, 1e-13;
 Matches 101: Conservative 69; Mismatches 169; Indels 82; Gaps 14;

QY 16 VFLEFLTWLSLKLLNLNRLFPQKDIYLYEYSLSTSPFVARNRTHYKDEVRKESVNSG 75
 Db 19 IFLLFAKALIRLLENPNQRIHPPTDLL-----DDL-QINCTN 56
 QY 76 IYQEPLEIKSLLEIRRDDIIDLEDDVYAMTSDCDIYQTLRGVAOKIVSKESFPPIAY 135
 Db 57 ILQGEKNE--NTLELINTKM---ENKMNNSIDRCQITLSMRFKAYLSEEEAFPLSEF 111
 QY 136 SLVERKDAIMVERLHAIFYNQHNICYIHYDRKAPTEKYAMNNLAKCESNFIASKLEAV 195
 Db 112 GLIVYKELSQVLFLLSSIIYQONECYIAGENSAPAFLLLEKELNCFPN---KRPI 166
 QY 196 EYAHISRLQADNLCLSLKSSIQMKRYVINLGGDFPLKSNFELYSCLKLGNAMLET 255
 Db 167 KWSGFELINSYVGCLEFLSHLSKDMKIFQYLSGVDPILKTNLEMYRIKLRLNG-----TV 221
 QY 256 KPNKSLERFTYHHLRVRVYEVYKLPRTNISKAPDPHNIQIFVGSAYFVYI 315
 Db 222 NIGIS-----TYEDRL-----LNGKNKTESP--LPLFKSSLSLIPRAANL 262
 QY 316 FNNSIYODFFAMSKDTPSPDEHFMT-----LIRVPGIPGEISRSADQVSDQSKTRLY 369
 Db 263 SSSSVPOQLLEFLRTTVADEGFWGLFGNKLDFNVPSFNNDPLTNMGWYVRHQL- 321
 QY 370 KW-----NYEGFFYPSCTGSHRSVCITYGAELFWLIKDGHWANKFDSKYDPIILK 423
 Db 322 -WVESECHNYMKD-----RS-CVFEGIDVYVNLKMSBALVAHKLYIESEBEAFPC 368
 QY 424 L 424
 Db 369 L 369

Search completed: May 2, 2002, 07:48:48
 Job time: 213 sec

Fri May 3 10:57:56 2002

us-09-645-192-2.ispt

Page 9